Discrete math exercises for foundation courses in computer science

Version 1.0 (Jun 22, 2026)

Alexandros V. Gerbessiotis

CS Department
NJIT
Newark, NJ 07102.

Email: alexg@njit.edu

©2023-2026. Alex. Gerbessiotis. All rights reserved.

Printed on June 22, 2026

DRAFT. Copyright (c) 2022-2026. Alex. Gerbessiotis.
All rights reserved. Not to be posted online
or on the web or to be made available outside of
copyright holder's web-page

Preface

This e-book contains a collection of approximately one thousand problems with indicative solutions on the subject that is known generically as discrete mathematics or foundations of computer science.

Thematically it is split into 12 areas that map to chapters. At the beginning of each chapter sometimes a brief overview of some definitions or results are described.

The e-book is intended primarily for sophomore to junior undergraduate students. High-school students can practice also with some of the material. Graduate students with limited exposure to discrete mathematics and computer science can benefit from it as well.

The last two chapters of number theory include in the form of problems most of the results that are usually described in the form of theorems and proposition in a elementary number theory textbook.

This material is neither final nor thoroughly proofread. It constitutes work in progress and might contain errors. It should be used in conjunction with other references if consulted for factual checking. Report discrepancies with other sources, or factual errors, or typos to the author. Distribution of this material in any form, without the expressly written or e-written consent of the author is not allowed.

©2022-2026. Alexandros Gerbessiotis. All rights reserved.

Contents

I Sets and Sequences **1**

1 Sets, relations, and functions **3**
1.1 Short review of sets, relations, and functions . . . 3
1.1.1 Sets . . . 3
1.1.2 Relations . . . 4
1.1.3 Orders . . . 6
1.1.4 Binary operators and operations . . . 7
1.1.5 Functions . . . 8
1.1.6 Numbers . . . 10
1.2 Problems on set operations . . . 14
1.3 Problems on relationships of set operations . . . 27
1.4 Inclusion exclusion principle . . . 37
1.5 Other problems on sets . . . 40
1.6 Problems on relations . . . 42
1.7 Problems on functions . . . 50
1.8 Numbers . . . 62

2 Boolean algebra and proofs **65**
2.1 Short review on boolean algebra and functions . . . 65
2.1.1 Boolean algebra . . . 65
2.1.2 Additional boolean functions . . . 65
2.1.3 Logical equivalence of functions. . . . 66
2.1.4 Laws of boolean functions . . . 66
2.1.5 Propositions . . . 66
2.1.6 Compound propositions and implications . . . 67
2.1.7 Conjunctive and disjunctive normal forms . . . 68
2.1.8 Boolean matrices . . . 68
2.1.9 Quantifiers . . . 69
2.1.10 De Morgan's Laws for Predicates . . . 69
2.1.11 Proof techniques . . . 70
2.1.12 Peano's axioms of Arithmetic . . . 71
2.1.13 Simplified Peano exposition . . . 71
2.1.14 Extensive Peano exposition . . . 72
2.1.15 Ordinary induction . . . 73
2.1.16 Strong induction . . . 73
2.1.17 Strong versus weak induction . . . 75
2.1.18 Different base case . . . 76
2.1.19 Why does induction work? . . . 76
2.2 Propositional calculus . . . 78

2.3 Predicate calculus . . . 87
2.4 Elementary proof techniques . . . 92
2.5 Peano-based proof methods: induction . . . 122
2.6 Miscellaneous problems . . . 135

3 Sequences **139**
3.1 Short review on sequences and their sums . . . 139
3.1.1 Finite sequence . . . 139
3.1.2 Infinite sequence . . . 139
3.1.3 Summation and product of terms of a sequence . . . 139
3.1.4 Series versus sum . . . 140
3.1.5 Arithmetic sum . . . 140
3.1.6 Recursive sequences . . . 140
3.1.7 Power series . . . 141
3.2 Infinite or finite sequences . . . 142
3.3 Fibonacci . . . 144
3.4 Binomial . . . 148
3.5 Arithmetic, geometric and cubic sums . . . 150
3.6 Harmonic sequence and sum . . . 155
3.7 Geometric sequence and sum . . . 159
3.8 Series . . . 168
3.9 Average seek time of a HDD . . . 180

4 Identities and Inequalities **183**
4.1 Short review on inequalities . . . 183
4.1.1 Quadratic polynomial . . . 183
4.1.2 Methods to show an inequality . . . 184
4.1.3 Convexity and concavity . . . 184
4.1.4 Function minima and maxima . . . 184
4.2 Simple inequalities . . . 187
4.3 Absolute value inequalities . . . 198
4.4 Factorial and combinatorials . . . 201
4.5 Combinatorial identity bounds . . . 208
4.6 Fundamental inequalities . . . 226
4.7 Natural/Mercator logarithm arithmetic . . . 253
4.8 Exponential and logarithmic inequalities . . . 269
4.9 Series . . . 280
4.10 Entropy and probability . . . 293
4.11 Maxima and Minima . . . 297

II Counting **303**

5 Combinatorics **305**
5.1 Short review on counting principles . . . 305
5.1.1 Addition principle . . . 305
5.1.2 Multiplication principle . . . 305
5.1.3 Generalizing the two principles . . . 305
5.1.4 Pigeonhole principle . . . 306
5.1.5 Permutations and k-permutations . . . 306
5.1.6 Sample k out of n without replacement . . . 307

5.1.7 Cyclic permutations . . . 307
5.1.8 Permutations with repetitions . . . 308
5.1.9 Derangements . . . 308
5.1.10 Counting subsets: k-of-n combinations . . . 308
5.1.11 Inclusion-Exclusion principle: two sets . . . 310
5.1.12 Inclusion-Exclusion principle: three sets . . . 311
5.1.13 Inclusion-Exclusion principle: more sets . . . 311
5.1.14 Partitions . . . 312
5.1.15 Selection and Multisets : combinations with repetitions . . . 313
5.1.16 Integer equation: number of non-negative solutions . . . 314
5.1.17 Integer equation: number of positive solutions . . . 314
5.1.18 Choosing k objects from n with repetitions: order matters . . . 314
5.1.19 Sample k out of n with replacement: order matters . . . 315
5.1.20 Choosing k out of n . . . 315
5.1.21 Partition identical objects into k possibly empty groups . . . 315
5.1.22 Partition identical objects into k non-empty groups . . . 316
5.1.23 Partition identical objects into k non-empty indistinguishable groups . . . 316
5.2 Addition and Multiplication principle: examples . . . 317
5.3 Pigeonhole examples . . . 339
5.4 Inclusion-exclusion: examples . . . 341
5.5 Permutations: examples . . . 350
5.6 Combinations: examples . . . 368
5.7 Ordered items : examples . . . 379
5.8 Combinatorial interpretation : examples . . . 384
5.9 Combinatorial identities : examples . . . 397
5.10 Grid examples . . . 407
5.11 Catalan number . . . 416
5.12 Derangements of permutations . . . 419
5.13 Towers of Hanoi . . . 424
5.14 Fibonacci related examples . . . 427
5.15 Miscellaneous examples . . . 431
5.16 Distributions . . . 462
5.17 Stirling's number of the second order . . . 467
5.18 Bell numbers . . . 471

6 Recurrences and generating functions 473
6.1 Short review on recurrences . . . 473
6.1.1 Arithmetic and geometric progressions . . . 473
6.1.2 Recurrence relations . . . 473
6.1.3 Recurrences of order k . . . 473
6.1.4 Linear Recurrences of order k . . . 474
6.1.5 Ordinary generating functions . . . 474
6.1.6 Exponential generating functions . . . 474
6.2 Aritmetic progressions . . . 475
6.3 Geometric progressions . . . 478
6.4 Generating functions . . . 480
6.5 Recurrences . . . 519

III Analysis of algorithms tools 549

7 Asymptotic notation and other recurrence relations 551

7.1 Short review on asymptotic notation . . . 551
7.1.1 Big-oh O notation . . . 551
7.1.2 Big-Omega Ω notation . . . 551
7.1.3 Theta Θ notation . . . 552
7.1.4 Little-oh o notation . . . 552
7.1.5 Little-omega ω notation . . . 552
7.2 Warm-up : limits . . . 554
7.3 Asymptotic notation . . . 557
7.4 Recurrence relation solution . . . 580
7.5 Recurrences and algorithm analysis: warming up . . . 645

IV Graphs 649

8 Introduction to graph theory 651

8.1 Short review on directed and undirected graphs . . . 651
8.1.1 Degree of a vertex . . . 651
8.1.2 Simple graphs . . . 652
8.1.3 Regular graphs . . . 652
8.1.4 Cycles, closed trails, tours, circuits in graphs . . . 652
8.1.5 Connectivity and subgraphs . . . 652
8.1.6 Trees . . . 653
8.1.7 Hamiltonian paths and cycles . . . 653
8.1.8 Matching in bipartite graphs . . . 653
8.1.9 Matchings in Graphs. . . . 653
8.1.10 Saturating vertices and sets . . . 654
8.1.11 Edge cover . . . 654
8.1.12 Vertex covers . . . 654
8.1.13 Alternating paths . . . 655
8.1.14 Augmenting paths . . . 655
8.1.15 Maximum, maximal and complete matchings . . . 655
8.1.16 Perfect matchings . . . 656
8.1.17 Neighborhood of a vertex set . . . 656
8.2 Elementary graph theory . . . 657
8.3 Trees, forests and connectivity in undirected graphs . . . 668
8.4 Eulerican cycles, paths and walks . . . 685
8.5 Hamiltonian cycles and paths . . . 703
8.6 Planar graphs . . . 719
8.7 Coloring on graphs . . . 733
8.8 Matchings . . . 751
8.9 Adjacency matrix and list . . . 789
8.10 Ramsey numbers . . . 792
8.11 Random Graphs . . . 806
8.12 Spanning trees . . . 808

V Probability **811**

9 Probability and random variables **813**

9.1 Short review on probability theory . . . 813
9.1.1 Experiments and trials . . . 813
9.2 Events . . . 814
9.2.1 Event space and measurable spaces . . . 814
9.2.2 σ-algebras . . . 814
9.2.3 Measurable spaces . . . 814
9.2.4 Topologies . . . 815
9.2.5 Measures . . . 815
9.3 Probability space . . . 816
9.3.1 Probability measure . . . 816
9.3.2 Probability space . . . 816
9.4 Probability of events . . . 818
9.5 Conditional Probabilities . . . 819
9.6 Random variables . . . 820
9.7 Discrete random variables . . . 821
9.7.1 Probability mass function (p.m.f.) . . . 821
9.7.2 Composition of random variables . . . 823
9.7.3 Expectation . . . 823
9.7.4 Joint probability . . . 826
9.7.5 Indepdendent random variables . . . 828
9.7.6 Variance . . . 828
9.7.7 Cumulative distribution function . . . 831
9.7.8 Probability generating function . . . 831
9.8 Continuous random variables . . . 833
9.8.1 Cumulative distribution function . . . 833
9.8.2 Probability density function (p.d.f.) . . . 834
9.8.3 Expectation of a continuous r.v. . . . 834
9.8.4 Independent random variables: continuous case . . . 834
9.8.5 Variance of a continuous r.v. . . . 835
9.8.6 Median and other quartiles . . . 836
9.8.7 Miscellanea . . . 837
9.9 Standard distributions and random variables . . . 838
9.9.1 Uniform . . . 838
9.9.2 Bernoulli distribution . . . 838
9.9.3 Binomial distribution . . . 838
9.9.4 Normal . . . 840
9.9.5 Geometric distribution . . . 841
9.10 Generating functions . . . 843
9.10.1 Probability generating function . . . 843
9.10.2 Moments . . . 843
9.10.3 Moment generating function . . . 843
9.10.4 Some examples . . . 845
9.11 Indicator random variable . . . 847
9.12 Characteristic function . . . 848
9.13 Conditional expectation . . . 849
9.14 Probability inequalities . . . 852
9.14.1 Markov's inequality . . . 853
9.14.2 Cauchy-Schwartz inequality . . . 855

9.14.3 Tsebyshev's or Tchebychef's inequality . . . 857
9.14.4 Jensen's inequality . . . 860
9.14.5 Holder's inequality . . . 861
9.14.6 Liapunov's inequality . . . 862
9.14.7 Minkowski's inequality . . . 863
9.14.8 Hoeffding inequality . . . 864
9.15 More on expectations . . . 866
9.16 Spaces . . . 868
9.17 Events and their properties . . . 872
9.18 Toys: dies, coins and card . . . 878
9.18.1 Throw or roll a die or dice . . . 879
9.18.2 Toss a coin . . . 883
9.18.3 Deck of cards . . . 894
9.19 Conditional probabilities . . . 897
9.20 Independence of events . . . 903
9.21 Discrete random variables . . . 919
9.22 Continuous random variables . . . 930
9.23 Generating functions . . . 940
9.24 Birthdays . . . 955
9.25 Miscellanea . . . 960
9.26 Binomial process and Bernoulli trials . . . 972
9.27 Chernoff and Hoeffding bounds . . . 975
9.28 Permutations and Derangements . . . 982
9.29 Selections and Bin and Balls . . . 989
9.29.1 ... or not . . . 993
9.30 Collecting coupons . . . 1008
9.31 Random graphs . . . 1011
9.32 Puzzles . . . 1027
9.33 Ramsey's numbers . . . 1029
9.34 Van der Waerden . . . 1035
9.35 Hashing . . . 1038
9.36 Miscellanea . . . 1047

VI Theory of numbers 1053

10 Introduction to Number theory 1055
10.1 Short review on number theory . . . 1055
10.1.1 Numbers . . . 1055
10.1.2 Binary operators and operations . . . 1056
10.1.3 Groupoids and semigroups . . . 1056
10.1.4 Groups . . . 1057
10.1.5 Rings and Integral domains . . . 1058
10.1.6 Fields . . . 1059
10.1.7 Divisibility . . . 1061
10.1.8 Congruences . . . 1062
10.1.9 Complete system of congruences . . . 1066
10.1.10 Reduced system of congruences . . . 1066
10.1.11 Congruence summary . . . 1067
10.1.12 Units . . . 1068
10.1.13 Inverses . . . 1069

10.2 Divisor Properties . . . 1071
10.3 Compositeness and Primeness . . . 1082
10.4 A theorem by Euclid . . . 1088
10.5 Integer Division . . . 1093
10.6 Greatest Common Divisor . . . 1100
10.7 Extended GCD . . . 1104
10.8 Euclid strikes again: GCD algorithm . . . 1115
10.9 Least Common Multiple . . . 1128
10.10 Fundamental Theorem of Arithmetic . . . 1134
10.10.1 Mersenne numbers . . . 1139
10.10.2 Fermat numbers . . . 1140
10.11 Factorization . . . 1144
10.12 Congruences . . . 1146
10.13 Diophantine equations . . . 1154
10.13.1 Linear congruences . . . 1158
10.13.2 Reduced system of congruences . . . 1162
10.14 Short review on primitive roots and quadratic residues . . . 1171
10.14.1 Euler's totient function . . . 1171
10.14.2 Units, zero divisors, quadratic residues . . . 1171
10.14.3 Legendre symbol . . . 1172
10.14.4 Order of a mod n . . . 1173
10.14.5 Primitive roots mod n . . . 1173
10.14.6 Jacobi symbol . . . 1174
10.15 Euler's totient function $\phi(n)$ . . . 1176
10.16 Modular equations . . . 1182
10.16.1 Wilson's theorem . . . 1185
10.16.2 Fermat's several theorems . . . 1188
10.16.3 RSA . . . 1198
10.17 Euler's theorem . . . 1200
10.18 Modular equations continued . . . 1204
10.18.1 Chinese remainder theorem . . . 1206
10.18.2 System of two equations . . . 1208
10.18.3 Generalized chinese remainder theorem . . . 1210
10.18.4 Modular system of polynomial equations . . . 1215
10.19 Diophantine equations continued . . . 1218
10.20 Primitive roots . . . 1225
10.20.1 Lagrange's theorem . . . 1244
10.20.2 Wilson's theorem revisited . . . 1250
10.20.3 Primitive roots continued . . . 1251
10.21 Legendre and Jacobi symbols . . . 1269
10.21.1 Summary of Legendre symbol properties . . . 1270
10.21.2 Summary of Jacobi symbol properties . . . 1272
10.21.3 Legendre symbol properties . . . 1275
10.21.4 Euler's criterion . . . 1280
10.21.5 More Legendre . . . 1285
10.21.6 Jacobi warmup . . . 1289
10.21.7 Jacobi symbol properties . . . 1291
10.21.8 Gauss lemma . . . 1296
10.21.9 Eisenstein theorem and more Legendre symbol properties . . . 1299

11 Primality testing **1319**
11.1 Carmichael numbers 1319
11.1.1 Korselt's theorem or criterion 1321
11.2 Fermat primality 1328
11.2.1 A Fermat Little Theorem-based primality test 1334
11.3 Solovay-Strassen primality test 1336
11.4 Probabilistic primality tests 1342
11.4.1 The Miller-Rabin primality test 1343
11.5 Prime number theorem 1366
11.5.1 Solovay-Strassen and Miller-Rabin failure probability bounds 1368

Part I

Sets and Sequences

Chapter 1

Sets, relations, and functions

1.1 Short review of sets, relations, and functions

1.1.1 Sets

Set. A set is a collection of elements in no particular order; an element can appear only once in a set.
Set notation. A set of elements is usually assigned a symbolic name. We can then refer to the set with its symbolic name. Thus set $\{10,30,20\}$ is assigned a symbolic name A as follows: we write $A = \{10,30,20\}$ then.
Writing a set; membership to a set. In set notation, for set $A = \{10,30,20\}$, we write $10 \in A$ to indicate that 10 is a member of set A, or equivalently, that 10 belongs-to set A. We also write $4 \notin A$ to indicate that 4 is not a member of set A.
Cardinality of a set. The cardinality of a set is the number of its elements.
The cardinality of A is three and we can write this as follows

$$|A| = 3, \quad \text{or} \quad \mathfrak{c}(A) = 3$$

Mapping elements of a set to natural numbers. Moreover we can map 10 to 1, 30 to 2, 20 to 3 thus we can map the elements of A to the subset $\{1,2,3\}$ of natural numbers, or the other way around we can map a subset $\{1,2,3\}$ of natural numbers to the elements of A.
Note about natural numbers. Sometimes (the set of) natural numbers will not include zero, sometimes it will include zero.
Empty set. A set with no elements is the empty set. Its cardinaly is 0. Set $B = \{\}$ is an empty set. It can also be denoted by $B = \emptyset$.
Subset of a set A. Set B is a subset of A if and only if every element of B belongs to A. It is a proper subset of A if A has at least one element not in B. We use the notation $B \subseteq A$ and $B \subset A$ respectively to denote a subset or proper subset. Therefore the following apply (using notation that has not been formally introduced/reviewed).

$$B \subseteq A \Leftrightarrow (\forall x)(x \in B \Rightarrow x \in A).$$

$$B \not\subseteq A \Leftrightarrow (\exists x)(x \in B \Rightarrow x \notin A).$$

$$B \subset A \Leftrightarrow (\forall x)(x \in B \Rightarrow x \in A) \wedge (\exists x)(x \in A \Rightarrow x \notin B).$$

Equal sets. Two sets are equal if they have the same elements. Then

$$A = B \Leftrightarrow (A \subseteq B) \wedge (B \subseteq A).$$

Comprehension of a set. If X is a set, and P defines a property of the elements of X, we denote the subset of X having property P by writing

$$\{x \in X : P(x)\}$$

An alternative for : is to use | or / or the symbol . which is the period.
Powerset of A. The powerset of A is the set containing all the subsets of A and it is denoted by $\mathbb{P}(A)$. Sometimes we write 2^A for $\mathbb{P}(A)$.
Universal set U. Universal set U is the set that contains all sets. That is, the following becomes true.

$$(\exists U)(\forall A)(A is\ set) \Rightarrow A \subseteq U.$$

Real numbers and intervals. A closed interval of real numbers, a closed/open and open/closed interval, and an open interval are defined as follows.

$$[x,y] = \{z \in \mathbb{R} : x \leq z \leq y\}.$$
$$[x,y) = \{z \in \mathbb{R} : x \leq z < y\}.$$
$$(x,y] = \{z \in \mathbb{R} : x < z \leq y\}.$$
$$(x,y) = \{z \in \mathbb{R} : x < z < y\}.$$

Union of sets. The union of two sets A,B is defined as follows

$$A \cup B = \{x : (x \in A) \vee (x \in B)\}.$$

Intersection of sets. The intersection of two sets A,B is defined as follows.

$$A \cap B = \{x : (x \in A) \wedge (x \in B)\}.$$

Difference of sets. The difference of two sets A,B is defined as follows.

$$A \setminus B = A - B = \{x : (x \in A) \wedge (x \notin B)\}.$$

Complement of a set. The complement $A^{\complement}$ (or A') of set A is defined as follows.

$$A^{\complement} = \{x : (x \notin A)\}.$$

Then we can define $A - B = A \cap B^{\complement}$.
Difference $U - A$ of sets. The difference $U - A$ is defined as

$$U - A = \{x : (x \in U) \wedge (x \notin A)\}.$$

Then we can define $A^{\complement} = U - A$.
Symmetric difference of two sets. The symmetric difference of two sets A,B is defined as

$$A \oplus B = (A - B) \cup (B - A).$$

Properties of sets.

Associative Law	$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$
Commutative Law	$A \cup B = B \cup A, \quad A \cap B = B \cap A.$
Distributive Law	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$
Identity	$A \cup \emptyset = \emptyset \cup A = A, \quad A \cap U = U \cap A = A.$
Inverse	$A^{\complement} \cup A = A \cup A^{\complement} = U, \quad A \cap A^{\complement} = A^{\complement} \cap A = \emptyset$
Double Complement	$(A^{\complement})^{\complement} = A.$
Dominance	$U \cup A = A \cup U = U, \quad \emptyset \cap A = A \cap \emptyset = \emptyset.$
Idempotence	$A \cup A = A, \quad A \cap A = A.$
De Morgan's Law	$(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}, \quad (A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$
Absorptive Law	$A \cup (A \cap B) = A, \quad A \cap (A \cup B) = A.$

1.1.2 Relations

Definition 1.1 **Binary relation**

A binary relation R on sets X and Y is a relation $R \subseteq X \times Y$ between elements of X and Y.

The domain of R is a subset of X. The codomain of R is Y. The range of R is a subset of the codomain.
How to describe a relation. We write xRy to indicate $x \in X, y \in Y$ are related, or $(x,y) \in R$. Also $(x,y) \notin R$ or $x \not R y$ to indicate that x and y are not related.

$$R = \{(x,y) \in X \times Y : xRy\}.$$

Definition 1.2 **Domain, codomain, and range**

The domain $d(R)$ of relation R is

$$d(R) = \{x | \exists y : xRy\}$$

The codomain of R is Y that is $cd(R) = Y$. The range of R is denoted by $r(R)$ and is

$$r(R) = \{y | \exists x : xRy\}$$

Thefore $r(R) \subseteq cd(R)$.

Definition 1.3 **Graph of a relation**

The graph G_R of relation R is defined as follows.

$$G_R = \{(x,y) \in X \times Y : xRy\}.$$

Definition 1.4 **On relation**

If $X = Y$ then relation R is a relation on X.

Definition 1.5 **Inverse of a relation**

For a relation R its inverse R^{-1} is defined as follows

$$R^{-1} = \{(y,x) | (x,y) \in R\},$$

or equivalently

$$R^{-1} = \{yR^{-1}x : xRy\}.$$

Definition 1.6 **Reflexive relation**

A binary relation R on X is a reflexive relation if for every $x \in X$ we have xRx.

Number of reflexive relations on set X. If X has n elements we can define $2^{n(n-1)}$ reflexive relations on them. Think of the graph of the relation with diagonal elements set to 1 and all other elements varying from 0 to 1. Let the number of elements lower than the diagonal of a square $n \times n$ matrix be x. Likewise over the diagonal is also x and the number of diagonal elements is n. The $x + x + n = n^2$, and thus $2x = n^2 - n$.

Definition 1.7 **Symmetric relation**

A binary relation R on X is a symmetric relation if xRy implies yRx.

Number of symmetric relations on set X. If X has n elements we can define $2^{n(n+1)/2}$ symmetric relations on them. Think of the graph of the relation. Its adjacency matrix must be symmetric in the sense $x_{ij} = x_{ji}$ indicating $x_i R x_j$ implies $x_j R x_i$. The on and over diagonal elements are $n(n+1)/2$.

Definition 1.8 **Antisymmetric relation**

A binary relation R on X, is an antisymmetric relation if xRy and yRx imply $x = y$.

Contrapositive of definition for antisymmetricity. Thus if $x \neq y$ then $x \not R y$, or $y \not R y$. (This is the contrapositive of the definition.) The last observation leads to the following. A relation is antisymmetric if there are no x, y such that xRy and yR. Thus if $R = \{(x,y)\}$, then R is antisymmetric by default.

Definition 1.9 **Transitive relation**

A binary relation R on X, is a transitive relation if xRy and yRz imply xRz.

1.1.3 Orders

Definition 1.10 **Types of relations**

A relation can be an order or an equivalence relation.

Definition 1.11 **Types of orders**

An order can be a partial order or a linear order also known as total order.

Definition 1.12 **Order**

A binary relation R on X is an order on X if for every $x, y, z \in X$ the following properties are satisfied.

- (a) Reflexivity, xRx for every $x \in X$,
- (b) antisymmetry xRy and yRx then $x = y$, and
- (c) transitivity xRy and yRz then xRz.

Sometimes we might write $a \preceq b$ for aRb. Then we can write $b \succeq a$. If $a \preceq b$ and $a \neq b$ then $a \prec b$. Replace $\prec$ and $\preceq$ with $<$ and $\leq$ and the similarity is then established.

Definition 1.13 **Poset**

The set X is a partially ordered set or poset and denoted by (X, R).

Definition 1.14 **Total/Linear order**

An order is linear (total) if either xRy or yRx.

Definition 1.15 **Partial order**

An order is a partial order if it is not a total order.

Example 1.1.1

$\leq$ is a total order on $\mathbb{Z}$ but / (divisibility) is a partial order.

Definition 1.16 **Lexicographic order**

Let R and S be two partial orders on X and Y. Then order L is defined on $X \times Y$ as follows.

$$(x_1, y_1) R (x_2, y_2)$$

if

$$x_1 R x_2 \vee x_1 = x_2 \wedge y_1 S y_2.$$

Definition 1.17 **Equivalence relation**

A binary relation R on X is an equivalence relation on X if for every $x, y, z \in X$ the following properties are satisfied.

- (a) reflexivity xRx for every $x \in X$,
- (b) symmetry xRy then yRx, and
- (c) transitivity xRy and yRz then xRz.

Definition 1.18 **Equivalence classes**

In an equivalence relation we can decompose X into disjoint subsets known as equivalence classes. Then for each $x \in X$ the set of elements equivalent to x define

$$[x/R] = \{y \in X : xRy\}.$$

Then

$$p : X \to X/R \text{ with } p(x) = [x].$$

For an equivalence relation R if aRb the $[a] = [b]$. (It is the same as $[a/R] = [b/R]$.)

1.1.4 Binary operators and operations

Definition 1.19 **Binary operator**

A binary operator defines a binary operation.

A binary operation is the operation implied by a binary operator.

Definition 1.20 **Binary operation**

For a set S and a binary operator $\oplus$, a binary operation is a rule that maps to each pair $(x, y) \in S \times S$ a unique $z \in S$.

Operations $+$, $*$ are binary operations for a set $S = \mathbb{N}$ or $S = \mathbb{R}$. In all definitions below we assume S is non-empty.

Definition 1.21

Operator $+$ for operation addition (and similarly $-$, $*$, etc) defines a binary operation.

Definition 1.22

Operator $<$ for operation comparison-less (and similarly $<$, $\leq$, $\geq$, $=$, $\neq$ etc) defines a binary relation that is a binary boolean function: it takes two arguments (numbers) a and b and returns a true (a 1) or false (a 0), as needed.

1.1.5 Functions

Definition 1.23 **Function**

A function $f : X \mapsto Y$ assigns to each $x \in X$ a unique element $f(x) \in Y$. X is the domain, and we write $d(f) = X$, and Y is the codomain of f and we write $\text{cd}(f) = Y$.

Definition 1.24 **Function: alternative definition**

A function $f : X \mapsto Y$ with domain X and codomain Y ($d(f) = X$ and $\text{cd}(f) = Y$ respectively) is a binary relation R between X and Y so that for every $x \in X$, there exists a unique $y \in X$ such that $(x,y) \in R$ (or xRy), and if xRy and xRz then $y = z$ (or equivalently $(x,y) \in R \wedge (x,z) \in R \rightarrow y = z$).

Every function is a relation. A relation may not be a function.

A function such as f defines a relation R on X and Y: xRy if an only if $xRf(x)$, i.e. $y = f(x)$. Thus functions are a special case of relations. For the relation R of function f with $yRf(x)$ or equivalently $(x,y) \in R$, there is only one pair (x,y) that contains x as its first element, and every $x \in X$ appears as a first element of a pair of the relation (function) f. For a general relation it is possible the same x is related with multiple y or some x is not related with any y at all.

Example 1.1.2

Let $f : X \mapsto Y$ and let $|X| > |Y|$. Then there are two elements $x_1, x_2 \in X$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2)$.

Definition 1.25 **Domain, Codomain, and Range or Image of f**

X is known as the domain $d(f) = X$ of f and Y is known as the codomain $\text{cd}(f) = Y$. The range or image of a function f denoted by $r(f)$ or $I(f)$ or just $image(f)$, is the set of values

$$r(f) = I(f) = image(f) = \{y \in Y : y = f(x) \text{for some } x \in X\}.$$

Note that we can also write $r(f) = r(X)$. Then $r(A)$, $A \subseteq X$ is defined $r(A) = \{f(x) : x \in A\}$.

Definition 1.26 **Graph of a function**

The graph G_f of function $f : X \rightarrow Y$ is the subset of $X \times Y$ defined as follows.

$$G_f = \{(x,y) \in X \times Y : x \in X, y = f(x)\}.$$

If f were a relation then G_f the graph of the relation defined as follows

$$G_f \subseteq \{(x,y) : x \in X, y \in Y\}, \quad , \quad \forall x \in X, \exists y \in Y : (x,y) \in f \quad , \quad (x,y) \in f \wedge (x,z) \in f \Rightarrow y = z,$$

classifies f as a functional relation i.e. a function.

The graph of a function f, i.e. G_f and the graph of a relation G_R might differ in two ways as noted earlie: there may be $x \in X$ such that $(x,y) \notin R$ for any $y \in Y$, and there may be $x \in X$ such that $(x,y) \in R$ for several $y \in Y$. Consider the case X is a group (set) of toddlers, and Y is a set of ice-cream flavors. You would be hard-pressed to find a toddler $x \in X$ who likes only flavor z and no other flavor! BTW the relation R implied by this function f is defined as the "likes" of the set of toddlers.

Definition 1.27 **Characteristic or indicator function of a set**

The characteristing or indicator function of set A sometimes denoted by χ_A is defined as follows.

$$A(x) = \chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Definition 1.28 **An into function**

A function f such that $f : X \mapsto Y$, is an into Y function. In other words we do not know if $\text{cd}(f) = r(f)$.

Definition 1.29 **An onto function is a surjectve function**

A function f such that $f : X \mapsto Y$, is an onto function if $cd(f) = r(f) = Y$. Such a function is also known as surjective.

A surjective function defines a surjection.

Definition 1.30 **An 1-1 function is an injective function**

A function f such that $f : X \mapsto Y$, is an one-to-one function if no member of Y is assigned to more than one member of X that is $x \neq y$ implies $f(x) \neq f(y)$. Such a function is also known as an injective function.

An injective function defines an injection. Distinct elements of X are mapped to distinct elements of Y.

Definition 1.31 **Bijective function**

A function f such that $f : X \mapsto Y$, that is both an surjection (onto) and injection (1-1) is a bijective function.

A bijective function defines an bijection.

Definition 1.32 **Partial function**

A partial function f assigns to each element of X at most one element of Y.

A partial function becomes a function when "at most one" becomes "exactly one" or "unique". Thus exactly one becomes at most one. Thus for $x \in X$ then either $f(x) \in Y$ or it is not defined (does not exist). Then the domain of f is the set of elements x such that $f(x) \in Y$. If the domain of f is X itself, then f is a total function.

Definition 1.33 **Total function**

A total function f assigns to each element of X exactly one element of Y.

Definition 1.34 **Variable, unknown, or indeterminate**

A function can be defined by way of a mathematical expression (formula). Thus $f(x) = x^2$ or $x \to x^2$, or $y = x^2$ map element x of $\mathbb{R}$ to element $f(x)$ or x^2 or y of $\mathbb{R}$. The x is known as a variable or indeterminate.

Definition 1.35 **Recursive functions**

A recursive function is a function that invokes itself. In direct recursion a recursive function f invokes directly itself, whereas in indirect recursion function f invokes function g that invokes f.

1.1.6 Numbers

Integer and natural numbers

Definition 1.36 **Natural numbers**

The set of natural numbers (non-negative integers) is denoted by $\mathbb{N}$.

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}.$$

We use a natural number to denote the cardinality of a set. Note that in an alternative definition of $\mathbb{N}$ number 0 is not included. The inclusion helps with set cardinality issues and several times subsequently with the use of induction.

Definition 1.37 **Integers**

The set of integers is denoted by $\mathbb{Z}$.

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}.$$

Definition 1.38 **Positive and negative integers**

The set of positive integers is denoted by $\mathbb{Z}_+$, the set of negative integers is denoted by $\mathbb{Z}_-$, and the set of non-zero integers is denoted by $\mathbb{Z}^*$.

$$\mathbb{Z}_+ = \{1, 2, 3, \ldots\}.$$
$$\mathbb{Z}_- = \{-1, -2, -3, \ldots\}.$$
$$\mathbb{Z}^* = \{\pm 1, \pm 2, \pm 3, \ldots\}.$$

For the positive and negative integers one may also use $\mathbb{Z}_+^*$ and $\mathbb{Z}_-^*$. $\mathbb{Z}$ is the set of integers (positive, negative or zero). For a positive integer $n > 0$ the set $\mathbb{Z}_n$ sometimes denoted by $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}/n$ is the set of integers modulo n, thus representing the n equivalence classes that the integers of $\mathbb{Z}$ can be split into depending on the remainder of their division by n.

Definition 1.39 **Rational numbers**

The set of rational numbers is denoted by $\mathbb{Q}$, where

$$\mathbb{Q} = \{\frac{a}{b} : a \in \mathbb{Z}, b \in \mathbb{Z}, b \neq 0\}.$$

Note that $\mathbb{Z} \subseteq \mathbb{Q}$.

Real numbers

Definition 1.40 **Real numbers**

The set of real numbers is denoted by $\mathbb{R}$.

Definition 1.41

The set of real numbers is denoted by $\mathbb{R}$. The set of non-zero real numbers will be denoted by $\mathbb{R}^*$. The set of positive real numbers will be denoted by $\mathbb{R}^*_+$. The set of negative real numbers will be denoted by $\mathbb{R}^*_-$. The set of non-negative real numbers will be denoted by $\mathbb{R}_+$. The set of non-negative real numbers will be denoted by $\mathbb{R}_-$.

Axioms of Real numbers

Axiom 1.1 **Axioms of real numbers**

A1. If $a \in \mathbb{R}$ then one and only one the following three statements is true.

A1.1. a is the unique number zero.

A1.2. a is a positive (real) number.

A1.3. $-a$ is a positive (real) number.

In the latter case, for $-a$ that is a positive number , we call real number a a negative number.

A2. If $a, b \in \mathbb{R}$ are positive, then $a+b$ is positive and $a \cdot b$ is positive.

A3. If $a \in \mathbb{R}$ is positive, then $1/a$ is positive.

The product of two numbers is denoted by the binary operator that is nothing or by $\cdot$ or by $*$.

Definition 1.42

Let $a, b \in \mathbb{R}$. The following are defined.

$$\begin{aligned} a > b &\Leftrightarrow a-b \text{ is positive number.} \\ a < b &\Leftrightarrow b > a. \\ a = b &\Leftrightarrow a-b \text{ is } 0. \end{aligned}$$

Equivalently $a < b \Leftrightarrow a-b$ is negative number.

Definition 1.43

We define $a \leq b$ to be a compound statement. It is true if either $a < b$ or $a = b$.

Definition 1.44

We define $a \geq b$ to be $b \leq a$.

Note that if a, b are real numbers, and a is positive and $-b$ is positive by Axiom A2, $a-b$ is positive. Then $a > b$. Note that if a, b are real numbers, and $-a$ is positive and b is positive by Axiom A2, $b-a$ is positive, or $a-b$ is negative. Then $b > a$.

Axiom 1.2

For any two real numbers $a, b \in \mathbb{R}$ one and only one of the following three statements is true.

B1. (Trichotomy property).

B1.1. $a = b$,

B1.2. $a > b$,

B1.3. $a < b$.

B2. For any $n \in \mathbb{N}$ we have $n \geq 0$. (The set of natural numbers includes 0 and positive integers.)

Theorem 1.1

The product of a positive number a with a negative number b is a negative number.

Proof. Let a be a positive number. Let b be a negative number. By Axiom [A1.3.] we have $-b$ a positive number. Then $a(-b) = -(ab)$ is also a positive number, which implies ab is a negative number. ■

Using once more the previous argument we derive the following.

Theorem 1.2

The product of a negative number a with a negative number b is a positive number.

As a consequence of the two previous theorems we have the following.

Theorem 1.3

For any real number a, the number $a^2 = a \cdot a$ is a positive number, or zero, if a is zero.

This is theorem part [C1.] below.

Theorem 1.4

Let $a, b, c, d \in \mathbb{R}$. The following are true.

C1. For all $a \in \mathbb{R}$, we have $a^2 \geq 0$.

C2. (Transitivity property) For any three real numbers $a, b, c \in \mathbb{R}$, if $a < b$ and $b < c$ then $a < c$.

C3. (Additivity property) For any three real numbers $a, b, c \in \mathbb{R}$, if $a < b$ and for any c then the following applies: $a + c < b + c$.

C4. (Multiply by positive) For any three real numbers $a, b, c \in \mathbb{R}$, and $c > 0$ if $a < b$ and for any $c > 0$ then the following applies: $ac < bc$.

C5. (Inequality add) If $a < b$ and $c < d$ then $a + c < b + d$.

C6. (Inequality pos-multiply) If $0 < a < b$ and $0 < c < d$ then $0 < ac < bd$.

C7. (Multiply by negative) If $a < b$ and $c < 0$ then $ac > bc$.

C8. (Inequality subtract) If $a < b$ and $c < d$ then $a - d < b - c$.

C9. (Positive inversion) If $0 < a < b$ then $1/a > 1/b$.

C10. (Inequality divide) If $0 < a < b$ and $0 < c < d$, then $\frac{a}{d} < \frac{b}{c}$.

C11. (Inequality exponentiate) If $0 < a < b$ then $a^n < b^n$ and $\sqrt[n]{a} < \sqrt[n]{b}$, for every integer $n \geq 2$.

C12. If $0 < a < b$ and $0 < c < 1 < d$, then $c^a > c^b$ and $d^a < d^b$.

Proof. Property [C3.] can be proved by the definition of $a < b$. If $a < b$ this means $b - a > 0$. If $b < c$ this means $c - b > 0$. Adding the two $c - a > 0$. For Property [C4.] we work as follows. If $a < b$ then $b - a$ is positive. Since c is also positive, by Axiom [A2] $c(b - a)$ is also positive. This yields $cb > ca$ or equivalently $ac < bc$ (use of commutativity and the definition for $a < b$ if and only if $b > a$. ■

Rational, Irrational, and algebraic numbers

Definition 1.45 **Rational and Irrational numbers**

A real number can be rational or irrational.

Definition 1.46 **Algebraic**

An algebraic number is the solution of a polynomial equation with integer coefficients.

All rational numbers are algebraic numbers as solution of $bx - a = 0$, $b \neq 0$, $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$. The only solution is $x = a/b$, a rational number.

Irrational numbers can be algebraic or transcedental. Real number $\sqrt{2}$ or $\sqrt{3}$ are irrational but algebraic as solutions of the polynomial equations $x^2 = 2$ and $x^2 = 3$ respectively. Real number π or e are irrational but not algebraic, and thus transcedental.

Transcedental numbers

Definition 1.47 **Transcedental**

A real number that is not algebraic is transcedental.

Definition 1.48 **Characterizing real numbers**

A real number can be rational or irrational. All rational numbers are algebraic. Irrational numbers are either algebraic or transendental.

Integer one is an integer, a rational and algebraic and thus a real number. Number $4/5$ is rational and algebraic and thus a real number. Number $\sqrt{2}$ is irrational, algebraic and thus a real number. Number π is irrational, transcendental and thus a real number.

1.2 Problems on set operations

Exercise 1.
(a) Find the cardinality of set C, where

$$C = \{\emptyset, 1, \{1\}\}.$$

(b) Does $1 \in C$? Does $\emptyset \in C$? Does $\{1\} \in C$?

Solution.
(a) Set C contains three elements: the empty set, integer 1, and a set containing integer 1. Its cardinality is 3 and therefore,

$$|C| = 3, \quad \text{or} \quad \mathfrak{c}(C) = 3.$$

(b) By way of part (a) YES is the answer to each one of the subquestions. ■

Exercise 2.
Let A,B,C are as follows: $A=\{1,2,3,4\}$, $B=\{1,3,5\}$, $C=\{2,4,6\}$. Find $A\cap B$, $A\cup B$ and $B\cap C$.

Solution.
For two sets A, B, $A\cap B$ denotes their intersection: $x\in A\cap B$ if and only if $x\in A$ and $x\in B$. For two sets A, B, $A\cup B$ denotes their union: $x\in A\cup B$ if and only if $x\in A$ or $x\in B$. The or is inclusinve in the sense that $x\in A$, or $x\in B$ or x is in both A and B. $A\cap B=\{1,3\}$. Moreover, $A\cup B=\{1,2,3,4,5\}$, and furthermore, $B\cap C=\emptyset$. ■

Exercise 3.
Let $A = \{2,3,5\}$, $B = \{2,4,6\}$, $C = \{1,3,5,7\}$. Find $A \cap B$, $B \cap C$, and $A \cap C$ and $A \cup B$, $B \cup C$, and $A \cup C$.

Solution.
$A \cap B = \{2\}$,
$B \cap C = \{\}$,
$A \cap C = \{3,5\}$.
$A \cup B = \{2,3,4,5,6\}$,
$B \cup C = \{1,2,3,4,5,6,7\}$,
$A \cup C = \{1,2,3,5,7\}$. ■

Exercise 4.
For set $D = \{2,4,6,8\}$ use a set comprehension to enumerate the elements of the set.

Solution.
If X is a set, and P defines a property of elements of X, we denote the subset of X having property P by writing

$$\{x \in X : P(x)\}$$

An alternative for : is to use | or / or the symbol . which is the period.
Then applying this to D we have $D = \{x \in \mathbb{Z} :$ **x is an even positive integer and** $x < 10\}$. ■

Exercise 5.
Let $U = \{1,2,3,4,5,6,7,8,9,10\}$. Let A, B, C be as follows: $A = \{2,3,5\}$, and $B = \{2,4,6\}$, and $C = \{1,3,5,7\}$. Find $A^{\complement}$, $A - B$, $A \oplus B$, $B \oplus C$.

Solution.
$A^{\complement} = \{1,4,6,7,8,9,10\}$, and $A - B = \{3,5\}$, and $A \oplus B = \{3,4,5,6\}$, and $B \oplus C = \{1,2,3,4,5,6,7\}$. ■

Exercise 6.
Show that $A = \{2,3,5\}$ is not a subset of $B = \{n \in \mathbb{N} : n \text{ is odd}\}$

Solution.
In order to show $A \not\subseteq B$ it suffices to find ONE element of A that is not in B. Let $x = 2$. $2 \in A$ but $2 \notin B$. This is because B contains odd numbers such as $1,3,5,7$ etc. Thus $A \not\subseteq B$. ■

Exercise 7.
Is $\{\emptyset, \{\}\}$ a set ?

Solution.
No. $\emptyset$ and $\{\}$ are the same elements. They are two different ways to say empty set. In a set every element is listed once. In our case the same element, the emptyset is included twice. The following two sets are sets indeed. Let $A = \{\emptyset\}$. Let $B = \{\{\}\}$. Moreover $A = B$ since $\emptyset = \{\}$ i.e. the two sets contain the same element. Moreover $c(A) = c(B) = 1$, the cardinality of each is one (element). ■

Exercise 8.

What is the Powerset $\mathbb{P}(A)$ of set $A = \{a, 1, 2\}$?

Solution.
$\mathbb{P}(A) = \{\{\}, \{1\}, \{2\}, \{a\}, \{1,2\}, \{1,a\}, \{2,a\}, \{1,2,a\}\}$. Note that $\mathfrak{c}(A) = 3$. The $\mathfrak{c}(\mathbb{P}(A)) = 8 = 2^3$. For each element x of A there are two possibilities: (a) x to be included in a set of $\mathbb{P}(A)$, and (b) x not to be included in a set of $\mathbb{P}(A)$. Given that we have three elements of A the total number of set possibilities is $2 \times 2 \times 2 = 8$.

```
1   2   a    Set
n   n   n    { }
n   n   y    { a }
n   y   n    { 2 }
y   n   n    { 1 }
n   y   y    { 2, a }
y   n   y    { 1, a }
y   y   n    { 1, 2 }
y   y   y    { 1, 2, a }
```

■

Exercise 9.
Which of the following is true?
(a)

$$\mathbb{P}(\{a,b\}) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$$

(b)

$$\mathbb{P}(a,b) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$$

(c)

$$\mathbb{P}(\{a,\{b\}\}) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$$

(d)

$$\mathbb{P}(\{a,\{b\}\}) = \{\emptyset, \{a\}, \{\{b\}\}, \{a,b\}\}$$

(e)

$$\mathbb{P}(\{a,\{b\}\}) = \{\emptyset, \{a\}, \{\{b\}\}, \{a,\{b\}\}\}$$

Solution.
(a) Correct.
(b) Incorrect. The a,b of $\mathbb{P}(a,b)$ is not a set. It should have been $\{a,b\}$ as in the first example.
(c) Incorrect. Element $\{b\}$ should have been included in $\mathbb{P}(\{a,\{b\}\})$ in the form $\{\{b\}\}$ as a set. Moreover identify also the issue with $\{a,b\}$.
(d) Incorrect. Identify the issue with $\{a,b\}$
(e) Correct. The issue with $\{a,b\}$ has been fixed! ■

Exercise 10.

Russell's paradox. Let

$$X = \{A : A \text{is a set and } A \notin A\}.$$

(a) Give two sets that are in X.
(b) Give a set that is not in X.

Solution.
(a) One such set is $\emptyset$. Moreover $\mathbb{Z}$, $\mathbb{Q}$ are also in X.
(b) Define set B recursively

$$B = \{ \overbrace{\{ \{\ldots\} \}}^{B} \}$$

ad infinitum. ■

Exercise 11.
For set $A = \{10, 20, 30\}$, establish (give, write, generate) $\mathbb{P}(A)$, the Powerset of A. What is the cardinality of the powerset of A? Note that $A \in \mathbb{P}(A)$.

Solution.
The powerset of A is the set containing all the subsets of A.

$$\mathbb{P}(A) = \{\emptyset, \{10\}, \{20\}, \{30\}, \{10, 20\}, \{10, 30\}, \{20, 30\}, \{10, 20, 30\}\}$$

The powerset of A has cardinality

$$|\mathbb{P}(A)| = 2^{|A|} = 2^3 = 8.$$

Sometimes we write 2^A for $\mathbb{P}(A)$. ■

Exercise 12.
Let $U = \{1,2,3,4,5,6,7,8,9\}$ and let $A = \{1,2,3,4\}$, $B = \{1,3,5\}$, $C = \{2,4,6\}$.
Give $A^{\complement}$, $A - B$, $A - C$, $A \oplus B$ (Note: $A \oplus B = (A - B) \cup (B - A)$.)

Solution.
$A^{\complement} = \{5,6,7,8,9\}$. $A - B = \{2,4\}$, $A - C = \{1,3\}$, $A \oplus B = \{2,4,5\}$. ■

1.3 Problems on relationships of set operations

Exercise 13.
(a) Show the following. For any two sets A,B, we have $A\cap B\subseteq A$, $A\cap B\subseteq B$.
(b) Then, show the following. For any two sets A,B, we have $A\subseteq A\cup B$, $B\subseteq A\cup B$.
(c) One can use Venn diagrams to visualize properties of sets. For $A,B\subset X$ we have the DeMorgan's Laws.

$$(A\cup B)^{\complement}=A^{\complement}\cap B^{\complement}.$$

and

$$(A\cap B)^{\complement}=A^{\complement}\cup B^{\complement}.$$

Prove the abovementioned assertions.

Solution.
(a) Consider set $C=A\cap B$. Let $c\in C$. Then since $C=A\cap B$, $c\in A$ and $c\in B$. The former implies $A\cap B\subseteq A$, and the latter $A\cap B\subseteq B$, since every element of c of C belongs to A and B respectively.
(b) It is proven similarly to part (a).
(c) We prove first $(A\cup B)^{\complement}\subseteq A^{\complement}\cap B^{\complement}$. Let $a\in(A\cup B)^{\complement}$. Then $a\notin(A\cup B)$. Thus $a\notin A, a\notin B$, or equivalently $a\in A^{\complement}, a\in B^{\complement}$, i.e. $a\in A^{\complement}\cap B^{\complement}$ thus proving $(A\cup B)^{\complement}\subseteq A^{\complement}\cap B^{\complement}$. The other ways is proven similarly and so does the other case. ■

Exercise 14.
For any two sets A, B, if $A \subseteq B$ and $B \subseteq A$ then $A = B$.

Solution.
Note that $A \subseteq B$ by definition means that for every $a \in A$ we have $a \in B$. In order to show that $A = B$ we need to show that for every $x \in A$ we have $x \in B$, and also for every $x \in B$ we have $x \in A$. We prove first the $x \in A \Rightarrow x \in B$. This is a direct consequence of $A \subseteq B$ which says $x \in A \Rightarrow x \in B$! We prove next the $x \in B \Rightarrow x \in A$. This is a direct consequence of $B \subseteq A$ which says $x \in B \Rightarrow x \in A$! Therefore A and B have the same elements and thus $A = B$. ■

Exercise 15.
For any two sets A, B, if $A = B$ then $A \subseteq B$ and $B \subseteq A$.

Solution.
This is the converse of the previous proposition. It can be proved similarly. ■

Exercise 16.
Let A,B,C be any three sets. Show that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

Solution.
(i) From $A \subseteq B$ we have $x \in A$ implies $x \in B$.
(ii) From $B \subseteq C$ we have $x \in B$ implies $x \in C$.
(iii) Thus by (i) every element $x \in A$ also belongs to B and thus $x \in B$.
(iv) Furthermore by (ii) every element in B belongs to C. Thus $x \in C$ as well.
Thus by way of (iii) and (iv) we have proved $x \in A$ implies $x \in C$ which means $A \subseteq C$. ■

Exercise 17.
Show that $X \cup (Y - X) = X \cup Y$.

Solution.
One way (the short way) is as follows. First $X \cup (Y - X) = X \cup (Y \cap X^{\complement})$. Then using the distributive law $X \cup (Y - X) = (X \cup Y) \cap (X \cup X^{\complement})$. The later part $X \cup X^{\complement} = U$ and thus $X \cup (Y - X) = X \cup Y$ as needed.
Another way (longer but surest) is by showing first $X \cup (Y - X) \subseteq X \cup Y$ and then $X \cup Y \subseteq X \cup (Y - X)$, implying that $X \cup (Y - X) = X \cup (Y \cap X^{\complement})$.
Consider an element z of $X \cup (Y - X)$. There are two reduced possibilities: (a) $z \in X$, or (b) $z \in Y - X$ and a third possibility, (c) $z \in X$ AND $z \in Y - X$ reduces to either (a), or (b), or (a) and (b).
(a) If $z \in X$ then by the definition of the union we have $z \in X \cup Y$.
(b) If on the other hand $z \in Y - X$ this means $z \in Y \cap X^c$ and thus $z \in Y$ in addition to being $z \in X^c$. Since $z \in Y$ then $z \in X \cup Y$ as well.
(c) For this case $z \in X$ and $z \in Y - X$ and thus by (a) or (b) $z \in X \cup Y$.
In all three cases $z \in X \cup (Y - X)$ leads to $z \in X \cup Y$ which implies $X \cup (Y - X) \subseteq X \cup Y$.
We now show $X \cup Y \subseteq X \cup (Y - X)$. Let $z \in X \cup Y$. We distinguish also three cases: (a) $z \in X$, and $z \notin Y$ (b) $z \in Y$, and $z \notin X$ (c) $z \in X, Y$ i.e. $z \in X \cap Y$. For case (a) and case (c) obviously $z \in X \cup (Y - X)$, by way of $z \in X$. For case (b), $z \in Y$ and since $z \notin X$ the latter implies $z \in X^{\complement}$. The $z \in Y - X$ and thus $z \in X \cup (Y - X)$. Thus in all three cases $X \cup Y \subseteq X \cup (Y - X)$. This establishes $X \cup Y = X \cup (Y - X)$. ■

Exercise 18.
Show that it is possible $A \cap B = A \cap C$ and yet $B \neq C$.

Solution.
Let $A = \{2,3\}$, $B = \{2,4,6\}$, $C = \{2,5,7\}$. It is clear that $A \cap B = A \cap C = \{2\}$, yet $B \neq C$. ■

Exercise 19.
For any A, B such $A \subseteq B$ we have $A \cap B = A$.

Solution.
In order to show $A \cap B = A$ (i) we first show $A \cap B \subseteq A$ and (ii)then show $A \subseteq A \cap B$.
(i) Let $x \in A \cap B$. This means $x \in A$ and $x \in B$. Because of the first part $x \in A$. Therefore we have shown that $x \in A \cap B \Rightarrow x \in A$. This means $A \cap B \subseteq A$.
(ii) Let $x \in A$. Let A, B are such that $A \subseteq B$, then $a \in A$ implies $a \in B$ for all a. Set $a = x$ and we have that $x \in A$ implies $x \in B$. Since $x \in A$ and $x \in B$ then $x \in A \cap B$. Therefore we have shown that $x \in A \Rightarrow x \in A \cap B$. This means $A \subseteq A \cap B$.
From $A \cap B \subseteq A$ and $A \subseteq A \cap B$ we have $A = A \cap B$. ■

Exercise 20.
Use properties of sets to compute in closed form $(A \cap B) \cup (A \cap B^{\complement})$. Closed form means the simplest way to describe a expression after all possible simplifications and reductions have been applied to it.

Solution.

$$\begin{aligned}
(A \cap B) \cup (A \cap B^{\complement}) &= ((A \cap B) \cup A) \cap \left((A \cap B) \cup B^{\complement}\right) \\
&= ((A \cup A) \cap (A \cup B)) \cap \left((A \cup B^{\complement}) \cap (B \cup B^{\complement})\right) \\
&= (A \cap (A \cup B)) \cap \left((A \cup B^{\complement}) \cap U\right) \\
&= (A \cup B) \cap (A \cup B^{\complement}) \\
&= (A \cup (B \cap B^{\complement})) \\
&= A \cup \emptyset = A
\end{aligned}$$

■

Exercise 21.
Find the number of 596-xyzw phone numbers that contain at least one 5 and at least one 6.

Solution.
Let A be the number of the given phone numbers that contain a 5. Their number is $|A| = 10^4 - 9^4$. Let B be the number of the given phone numbers that contain a 6. Their number is $|B| = 10^4 - 9^4$. In both cases 10^4 is the number of all phone numbers 596-xyzw as we have 10 choices for x, y, z, w and 9^4 is the number of phone numbers that only 9 of the 10 digits participate (either 5 or 6 is missing). The union $A \cap B$ is the numbers that contain at least one 5 and at least one 6. Thus

$$|A \cup B| = 10^4 - 8^4$$

where 10^4 is the number of all phone numbers 596-xyzw as we have 10 choices for x, y, z, w, and 8^4 is those numbers that have neither a 5 nor 6 in them, thus having the remaining 8 digits only. The intersection $A \cap B$ is the number of phone numbers with at least a 5 and at least a 6. We have

$$|A \cup B| = |A| + |B| - |A \cap B| = (10^4 - 9^4) + (10^4 - 9^4) - (10^4 - 8^4) = 10^4 - 2 \cdot 9^4 + 8^4 = 974.$$

■

1.4 Inclusion exclusion principle

Exercise 22.

The Cartesian product $X \times Y$ is the set of ordered pairs (x,y) with $x \in X$ and $y \in Y$.
Prove the Theorem of Inclusion-Exclusion for two sets A, B.
Prove the Theorem of Inclusion-Exclusion for three sets A, B, C by using the previous result for $A \cup B$ and C.

Solution.
Let us restate the theorem of inclusion exclusion for two sets A and B as follows.

$$c(A \cup B) = c(A) + c(B) - c(A \cap B).$$

If we have a count $c(A)$ of the elements of A and a count $c(B)$ of the elements of B, the count of $c(A \cup B)$ is not (necessarily) equal to the count $c(A) + c(B)$ since the latter count the elements of the intersection $A \cap B$ twice, once as elements of A and once as elements of B. Thus in order to calculate $c(A \cup B)$ first we calculate the sum $c(A \cup B)$ by double counting each element of $A \cap B$ and then subtracting from the sum $c(A) + c(B)$ $c(A \cap B)$ to correct the double counting.
Another way to say this is as follows. Let c belongs to ONE of the two sets and not the other; let without loss of generality $c \in A$ and thus $c \notin B$. Element c is counted through $c(A)$ but neither through $c(B)$ nor $c(A \cap B)$ as it does not belong to B. Thus it is counted only once, as it is supposed to be. Likewise if c is in B and not in A.
Let now c belong to A and to B and thus to $A \cap B$. Element c is counted once as belonging to A through $c(A)$, and once more as belonging to B through $c(B)$; need to adjust this double counting by $-c(A \cap B)$ as it subtracts one for every element of $A \cap B$ that includes c. ■

Exercise 23.

There is a group of students preparing to take CS435, whose prerequisites are CS241 and CS288. In this group, there are 30 students who have taken CS241, and there are 50 students who have taken CS288, and only 20 students who have taken both CS241 and CS288. What is the total number of students who have completed CS241 or CS288?

Solution.
Let A be the set of CS241 takers, Let B be the set of CS288 takers, Let $A \cup B$ be the groupd of students involved, and let $A \cap B$ be the 20 students among them who have taken both CS 241 and CS288. $|A| = 30, |B| = 50, |A \cap B| = 20$. Then by the inclusion exclusion principle, we have that

$$|A \cup B| = |A| + |B| - |A \cap B|. = 30 + 50 - 20 = 60.$$

■

1.5 Other problems on sets

Exercise 24.

Let set A contain n distinct objects, say the integers from 1 through n. Assume $n > 4$. Let set B contain all the subsets of A of cardinality 4. Find the cardinality of set B.

Solution.

Let D be an element of B. D contains four elements of A. How many different D we have is equal to how many ways can we construct D. Let the elements of D be a,b,c,d. We can pick a in n different ways from A. The choices left to choose for the next element of D and let it be b, are then limited to $A-\{a\}$. This is because a,b,c,d must be distinct as in a set we have no repetitions of an element. Then for b we can pick it in $n-1$ different ways from A as we cannot pick a again, in other words b is chosen from $A-\{a\}$ and there are $n-1$ elements in the latter. Likewise for c we have $n-2$ choices to choose from from $A-\{a,b\}$, and for d we have $n-3$ choices. Thus so far we have $n\cdot(n-1)\cdot(n-2)\cdot(n-3)$ choices to pick a,b,c,d. But in a set the order of the items does not matter. So if we picked $a=1$, $b=4$, $c=6$ and $d=7$ (assuming $n>7$) this choice forms the same set as the one determined by the choice say $a=6$, $b=4$, $c=1$ and $d=7$. Therefore all $4\cdot3\cdot2\cdot1=24$ rearrangements of the chosen values of a,b,c,d can lead to the same set. Thus for $1,4,6,7$ the total 24 rearrangements (including the original one) are shown below and are known as permutations of $1,4,6,7$.

$(1,4,6,7)$	$(1,4,7,6)$	$(1,6,4,7)$	$(1,6,7,4)$	$(1,7,4,6)$	$(1,7,6,4)$
$(4,1,6,7)$	$(4,1,7,6)$	$(4,6,1,7)$	$(4,6,7,1)$	$(4,7,1,6)$	$(4,7,6,1)$
$(6,1,4,7)$	$(6,1,7,4)$	$(6,4,1,7)$	$(6,4,7,1)$	$(6,7,1,4)$	$(6,7,4,1)$
$(7,1,4,6)$	$(7,1,6,4)$	$(7,4,1,6)$	$(7,4,6,1)$	$(7,6,1,4)$	$(7,6,4,1)$

Therefore the cardinality of D is

$$\frac{n\cdot(n-1)\cdot(n-2)\cdot(n-3)}{4\cdot3\cdot2\cdot1}.$$

We denote it as $\binom{n}{4}=\frac{n!}{(n-4)!\cdot4!}$. $n!$ that reads as n factorial is the product of the first n natural integers $1,2,3,\ldots,n$ and thus formally defined as

$$n!=n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot2\cdot1.$$

■

1.6 Problems on relations

Exercise 25.
Let $A = \{1,2,3\}$ and let $R = \{(1,1),(2,2),(3,3),(1,2),(2,1)\}$. Is relation R on A reflexive ? Is it symmetric ? Is it transitive ? Is it an equivalence relation?

Solution.
It is reflexive since $(1,1),(2,2),(3,3) \in R$. It is symmetric since $(a,b) \in R$ implies $(b,a) \in R$. And it is transitive; it satisfies all the conditions that define R as an equivalence relation. Moreover $[1] = \{1,2\}$, $[2] = \{1,2\}$, and $[3] = \{3\}$. Moreover $A/R = \{[1],[3]\}$ although $A/R = \{[2],[3]\}$ could also have been used since $[1] = [2] = \{1,2\}$. ■

Exercise 26.
Let $A = \{10011, 1001, 1110, 11011\}$ Let $B = \{2,3,4,5,6\}$.
(a) What is the cardinality of $A \times B$?
(b) Is $T = \{(1001,2)(1111,3),(11011,5)\}$ a relation from A to B?
(c) How many relations can we have from A to B.
(d) Is $R = \{(1001,2)(1110,3),(10011,3),(11011,4)\}$ a relation from A to B?
(e) For R find $d(R)$ and $r(R)$.
(f) For R find R^{-1}.
(g) For R find $d(R^{-1})$ and $r(R^{-1})$.

Solution.
(a) $c(A \times B) = c(A) \cdot c(B) = 5 \cdot 4 = 20$.
(b) No it is not. Element 1111 of $(1111,3)$ is not a member of A and thus $(1111,3) \not\subseteq A \times B$. For T to be a relation $T \subseteq A \times B$.
(c) There are $c(A) \cdot c(B) = 4 \cdot 5 = 20$ pairs in $A \times B$. Any subset of $A \times B$ is a relation. There are $2^{20} = 1,048,576$ candidates for a relation.
(d) Yes it is. Observe by the way for $(a,b) \in A \times B$, b is the number of ones of the binary number indicated by a (or just the number of ones of a if you know nothing about binary numbers).
(e) $d(R) = A$, $r(R) = B - \{5,6\} = \{2,3,4\}$
(f) $R^{-1} = \{(2,1001)(3,1110),(3,10011),(4,11011)\}$.
(g) $r(R^{-1}) = d(R) = A$ and $d(R^{-1}) = r(R) = \{2,3,4\}$. ■

Exercise 27.

Show that the relation R_3 defined on $\mathbb{Z}$ (the set of integers) is an equivalence relation. Relation R_3 i.e. xR_3y is defined as follows.

$$xR_3y \Leftrightarrow x \equiv y \pmod 3$$

That is x and y are related if and only if the difference $x-y$ or $y-x$ is a multiple of 3 or in other words 3 divides the difference $x-y$ (or $y-x$). We can also read the relationship as a congruence modulo three.

Solution.
Given relation $R_3 : x \equiv y \pmod 3$ defined on $\mathbb{Z}$, we will show it is reflexive, symmetric and transitive. $x \equiv y \pmod 3$ means $x-y$ is a multiple of 3 or equivalently 3 divides $x-y$ (and also $y-x$). That is there exists an integer k such that $x-y=3k$.
Reflexive. Obviously $x-x=3\cdot 0$. Thus 3 divides $x-x$ and thus $x \equiv x \pmod 3$. Therefore xR_3x.
Symmetric. If $x \equiv y \pmod 3$ there exists k such that $x-y=3k$. Then $y-x=3(-k)$ and thus $y \equiv x \pmod 3$ as well. Therefore xR_3y implies yR_3x.
Transitivity. Let xR_3y and yR_3z for some $x,y,z \in \mathbb{Z}$. Then $x \equiv y \pmod 3$ and $y \equiv z \pmod 3$. These imply $x-y=3k$ and $y-z=3m$ for some integer k,m. Then $(x-y)-(y-z)=3(k-m)$ for some integer $k-m$. But the left hand side is $x-z$ and thus $x-z=3(k-m)$. We conclude xR_3z.

■

Exercise 28.
Consider $\mathbb{Z}$. We say x and y are equivalent modulo 7 if

$$x \equiv y \pmod 7$$

i.e. $x-y$ is a multiple of 7, or equivalently, 7 divides $x-y$.
Show that the $\equiv$ relation is reflexive, symmetric and transitive.

Solution.
(a) Reflexivity. For every $a \in \mathbb{Z}$, we have $a-a=0\cdot 7$, thus 7 divides $(a-a)$. Thus $a \equiv a \pmod 7$.
(b) Symmetricity. Say $a \equiv b \pmod 7$. Then 7 divides $a-b$ or equivalently $a-b=7k$ for some integer k (i.e. $a-b$ is a multiple of 7. The $b-a=7(-k)$. If k is integer so is $-k$. Thus $b \equiv a \pmod 7$.
(c) Reflexivity. If $a \equiv b \pmod 7$, $b \equiv c \pmod 7$, then from the former we have $a-b=7k$ and similarly from the latter $b-c=7m$ for some integer k,m. The $a-c=(a-b)+(b-c)=7(k+m)$ with $k+m$ integer Thus $a \equiv c \pmod 7$. ■

Exercise 29.
Consider set $\mathbb{Z}$ (the set of integers) and relation R_7 (congruence modulo 7) i.e. xR_7y if and only if

$$x \equiv y \pmod 7$$

in other words x are y are related if their difference is a multiple of 7 or in other words 7 divides the difference $x-y$ (or $y-x$). Is R_7 an equivalence relation? What are its equivalence classes?

Solution.
R_7 is an equivalence relation (easy to show, see previous question) and the equivalence classes are as follows.

$$E_0 = \{\ldots,-21,-14,-7,0,7,14,21,\ldots\}$$

$$E_1 = \{\ldots,-20,-13,-6,1,8,15,22,\ldots\}$$

$$E_2 = \{\ldots,-19,-12,-5,2,9,16,23,\ldots\}$$

$$E_3 = \{\ldots,-18,-11,-4,3,10,17,24,\ldots\}$$

$$E_4 = \{\ldots,-17,-10,-3,4,11,18,25,\ldots\}$$

$$E_5 = \{\ldots,-16,-9,-2,5,12,19,26,\ldots\}$$

$$E_6 = \{\ldots,-15,-8,-1,6,13,20,27,\ldots\}$$

■

Exercise 30.
Consider set $A = \{a,b,c\}$ and relation on $A \times A$, such that $R = \{(a,a),(b,b),(a,c),(a,b)\}$.
Is R (a) reflexive, (b) symmetric, (c) antisymmetric, and (d) transitive?

Solution.
(a) NO. For $c \in A$, $(c,c) \notin R$.
(b) NO. $(a,c) \in R$ but $(c,a) \notin R$.
(c) YES. $(a,a) \in R$ and thus $a = a$ Likewise for (b,b). Even if there are $(a,c),(a,b)$ there are no (c,a) nor (b,a).
(d) YES. $(a,a),(a,c)$ imply (a,c). $(a,a),(a,b)$ imply (a,b). $(a,b),(b,b)$ imply (a,b). ■

Exercise 31.

Consider a relation where X are all n-bit words (strings). Let the elements of X be x_i, where $i = 0, \ldots, 2^n - 1$, and $N = 2^n$. There are 2^n elements in X. The question is whether there exists a relation R such that $z_i R z_{i+1}$ and $z_{N-1} R z_0$ and z_i is a reordering of x_i, where related elements are strings that differ in exactly one bit. We call such z_i a Gray code of X.
Show that there is a Gray code for every $n \geq 1$.

Solution.
We can construct a Gray code, by composition, with a method that will later be described as inductive construction. For one bit strings, $x_0 = 0$ and $x_1 = 1$ the relation has $y_0 R y_1$ where $x_0 = y_0$ and $x_1 = y_1$. For two bit strings $x_0 = 00, x_1 = 01, x_2 = 10 and x_3 = 11$. Then $y_0 = 00, y_1 = 01, y_2 = 11, y_3 = 10$, i.e. $y_0 = x_0, y_1 = x_1, y_2 = x_3, y_3 = x_2$. Suppose we have constructed a Gray code for n bit strings and it is $y_0, \ldots, y_{N-1}$. We construct a Gray code for $n+1$ bit strings as follows.

$$0y_0, 0y_1, \ldots, 0y_{N-2}, 0y_{N-1}, 1y_{N-1}, 1y_{N-2}, \ldots, 1y_1, 1y_0$$

Whether it is $y_0, y_1, \ldots, y_{N-2}, y_{N-1}$ or drawn the other way around $y_{N-1}, y_{N-2}, \ldots, y_1, y_0$ it is a gray code. We put those two back to back but the first set's code is preceded by a 0 and the latter set's Gray code with a 1. The $0y_{N-1} R 1y_{N-1}$ is indeed so as the two $n+1$-bit strings differ in only the left most bit. ■

1.7 Problems on functions

Exercise 32.
(a) For $f : R \longrightarrow R$ show that the function $f(x) = x^2$ is not injective. Moreover, what is X? What is Y? What is $I(f)$?
(b) Let $g : Y \longrightarrow Y$, Y is finite and g is a function. The g is 1-1 if and only if g is ONTO.

Solution.
Implicit is the definition $A = B = \mathbb{R}$. But $I(f) = \mathbb{R}_+$.
The function is not injective. This is because for x and $-x$ we have $f(x) = f(-x) = x^2$. Thus for a given $x^2 \in R$ there exist two elements of R, x and $-x$ mapping to x^2. On the other hand For $f : R_+ \longrightarrow R_+$ the function $f(x) = x^2$ is injective.
If Y that is finite has $c(Y)$ element and g is 1-1 the range of Y has $c(Y)$ elements and thus it is $r(g) \subseteq Y$. Since $r(g)$ and Y have the same cardinality $c(Y)$ they are equal to each other and $r(g) = Y$ implies g is ONTO.
If g is onto but say not 1-1 then some paire of Y, say $a, b \in Y$ get mapped to the same element i.e. $g(a) = g(b)$. Then if the remaining $n-2$ elements of Y get mapped by g to distinct elements the sum of those $n-2$ elements plus $g(a)$ give at most $n-1$ elements and thus one element of Y, say z is such that there is no $y \in Y$ such that $g(y) = z$. But this implies g is not ONTO, that leads to a contradiction to the fact that g is ONTO! ■

Exercise 33.

The composition of functions $f : X \to Y$ and $g : Y \to Z$, is denoted by $g \circ f : X \to Z$ and defined by

$$(g \circ f)(x) = g(f(x)).$$

An 1-1 and ONTO function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ defined as follows

$$f^{-1}(y) = x \text{ iff } f(x) = y.$$

Let $f(x) = 1/x^2$. Show that it is a function from the nonzero reals to the set of positive reals.

Solution.
Relation f is a function. For every real $x \neq 0$, x^2 is also a real and so is $1/x^2$. We can't have two different y_1, y_2 such that $y_1 = 1/x^2$ and $y_2 = 1/x^2$ as this would imply $y_1 = y_2$.
Function f is not one-to-one (injective). This is because for $x \neq 0$, both x and $(-x)$ map to the same $1/x^2$.
Function f is onto (surjective). This is because for every y that is positive we can find $y = f(x) = 1/x^2$ i.e. $x = \pm 1/\sqrt{y}$. Pick one of the possible x values, say $x = 1/\sqrt{y}$. We find $f(1/\sqrt{y}) = y$ and thus every y has an (at least one) x mapping to it. ■

Exercise 34.

Let $f : X \to \mathbb{R}$, $g : Y \to \mathbb{R}$. Then

$$(f+g)(x) = f(x)+g(x) \quad (f-g)(x) = f(x)-g(x) \quad (fg)(x) = f(x)\cdot g(x)$$

The functions are defined wherever both f, g are defined i.e. for example $dom(f+g) = dom(f) \cap dom(g)$.
A polynomial function $f : \mathbb{R} \to \mathbb{R}$ is defined as $f(x) = a_0 + a_1 x + \ldots + a^i x^i + \ldots a_n x^n$, where $n \neq 0$. The parameter n the degree of f and $n+1$ is the degree bound of f. We use $deg(f)$ and $db(f)$ to denote them respectively.
If $n = 1$ then f is a linear function, if $n = 2$ then it is a quadratic function and if $n = 3$ it is a cubic function.
$f(x) = b^x$ is an exponential function with $f : \mathbb{R} \to \mathbb{R}+$, $b \in \min\{R+\}$, $b \neq 1$ where $\mathbb{R}+$ denotes the positive reals, and $\mathbb{R}-$ would denote the negative reals. Parameter b is the base. For $y = b^x$ we call x the logarithm base b of y and it is denoted by $x = \log_b x$.
$f(y) = \log_b y$ is a logarithmic function with $f : \mathbb{R}+ \to \mathbb{R}$, $b \in \min\{R+\}$, $b \neq 1$ where $\mathbb{R}+$ denotes the positive reals, and $\mathbb{R}-$ would denote the negative reals. Parameter b is the base of the logarithms. If $b = 2$ we can use $\lg y$ instead of $\log_2 y$. If $b = 10$ we can use $\log y$ instead of $\log_{10} y$. If $b = e$ we can use $\ln y$ instead of $\log_e y$, where e is the base of the Neperian logarithms (to be defined later).
$f(x) = |x|$ is the absolute value function, $f : \mathbb{R} \to \mathbb{R}+$, where for $x \geq 0$, $f(x) = |x| = x$ and for $x < 0$ $f(x) = |x| = -x$.
$f(x) = \lfloor x \rfloor$, $g(x) = \lceil x \rceil$, are defined as the largest integer $\leq x$, and the smallest integer $\geq x$ respectively.
(a) Find $\lfloor 3.5 \rfloor, \lfloor 3 \rfloor, \lfloor -3.5 \rfloor, \lceil 3.5 \rceil, \lceil 3 \rceil, \lceil -3.5 \rceil$.
(b) Let $a, b, a > b$ be positive integers. The number of integers that are $\leq a$ and divisible by b is $\lfloor (a/b) \rfloor$.
(c) Find the number of (trailing i.e. ending) zeroes in 100 factorial $100! = 100 \cdot 99 \cdot 98 \cdot \ldots \cdot 2 \cdot 1$.

Solution.
$\lfloor 3.5 \rfloor = 3, \lfloor 3 \rfloor = 3, \lfloor -3.5 \rfloor = -4$, $\lceil 3.5 \rceil = 4, \lceil 3 \rceil = 3, \lceil -3.5 \rceil = -3$.
The positive integers divisible by b are multiples of b such that $b, 2b, 3b, \ldots, ib$ with $ib \leq a$. Moreover $(i+1)b > a$. Therefore $i \leq a/b$ and $i > a/b - 1$. The only integer in the range is a/b or $\lfloor (a/b) \rfloor$. In the former case $a/b = \lfloor (a/b) \rfloor$ and the result follows.
There are $100/2$ multiples of 2 in the integers $1, \ldots, 100$. There are $100/5$ multiples of 5 in the integers $1, \ldots, 100$. There are $100/25$ multiples of 25 in the integers $1, \ldots, 100$. Therefore, 5 divides $100!$, $20+4$ times or in other words 5^{24} divides $100!$. For each such five there a factor of 2 dividing $100!$ (there are $100/2 = 50$ of them overall). Thus there are 24 factors of $5 \cdot 2 = 10$ in $100!$ and the latter has 24 trailing zeroes. ■

Exercise 35.

A year y is a leap year if it is a centennial year divided by 400 or if it is a non-centennial year it is divisible by 4. Find the number of leap years after 1752 to year y, where $y > 2024$.

Solution.

Let r be a year such that $1752 < r \leq y$. The number of leaps years will be denoted by l. Let $4a$ be a year. Then $1752 < 4a \leq y$. Thus $1752/4 < a \leq y/4$. The number of leap years is thus $\lfloor (y/4) \rfloor - 1752/4 = \lfloor (y/4) \rfloor - 438$. Let $100c$ be a centennial year. Then we have $1752 < 100c \leq y$, The number of leap centennial years is thus $\lfloor (y/100) \rfloor - \lfloor (1752/100) \rfloor = \lfloor (y/100) \rfloor - 17$. Let $400d$ be a quad-centennial year. Then we have $1752 < 400c \leq y$, The number of leap centennial years is thus $\lfloor (y/400) \rfloor - \lfloor (1752/400) \rfloor = \lfloor (y/400) \rfloor - 4$. Therefore $l = a - c + d = \lfloor (y/4) \rfloor - 438 - \lfloor (y/100) \rfloor + 17 + \lfloor (y/400) \rfloor - 4 = \lfloor (y/4) \rfloor - \lfloor (y/100) \rfloor + \lfloor (y/400) \rfloor - 425$. ■

Exercise 36.
Let $A = \{10011, 1001, 1110, 11011\}$ Let $B = \{2,3,4,5,6\}$.
(a) Is $R = \{(1001,2),(1110,3),(10011,3),(11011,4)\}$ a relation from A to B?
(b) Is $R : A \to B$ as defined in (a) a function from A to B?
(c) Is T a function from A to B where $T = \{(1001,2),(10011,3),(11011,4)\}$

Solution.
(a) YES it is, from the discussion of the previous Chapter and a relevant exercise.
(b) YET it is. For all $a \in A$ there is a $b \in B$ such that aRb. This is function R definition. Thus the relation R is a function.
(c) NO it is not. There is an $a \in A$ for which there is no $b \in B$ such that aRb. This a is $a = 1110$. ■

Exercise 37.
For $f : R \to R$ and $g : R \to R$ we define $f(x) = 3x+1$ and $g(x) = x^2 - 4$. Find $g \circ f$.

Solution.
Note that $(g \circ f)(x) = g(f(x)) = g(3x+1) = (3x+1)^2 - 4 = 9x^2 + 6x + 1 - 4 = 9x^2 + 6x - 3$.
Moreover $f(x) = 3x+1$ shows $y = 3x+1$. And $f(x) = x^2 - 4$ implies $z = f(y) = y^2 - 4$. Eliminating y we have $z = y^2 - 4 = (3x+1)^2 - 4 = 9x^2 + 6x - 3$. ■

Exercise 38.

Two sets are equivalent if their cardinalities are the same and we write $X \equiv Y$. Then $\equiv$ is an equivalence relation on sets.

A set X is finite if $X \equiv \{1,2,\ldots,n\}$ for some fixed $n \in \mathbb{N}$.

A set X is countably infinite if $X \equiv \mathbb{N}$.

A set X is countable if it is countably infinite or finite.

A set X is uncountable if it is not countable.

(a) Show that $A = \{1,9,27,81,241,\ldots\}$ is countably infinite.
(b) The cartesian product $\mathbb{N} \times \mathbb{N}$ is countably infinite.
(c) The set of integers $\mathbb{Z}$ is countably infinite.
(d) A set X that is not empty is countable if and only if there is a onto map $f : \mathbb{N} \to X$.
(e) A countable union of countable sets is countable.
(f) The powerset $\mathbb{P}(\mathbb{N})$ of $\mathbb{N}$ is uncountable.

Solution.
(a) Define a function $f : \mathbb{N} \to A$ defined by $f(n) = n^3$. It is 1-1 and ONTO.
(b) Define a diagonalization i.e. a linear order $\ll$ where $(a,b) \ll (c,d)$ if either $a+b < c+d$ or $a+b = c+d$ and $b < d$. The listing of some pairs is

$$(1,1),(2,1),(1,2),(3,1),(2,2),(1,3),\ldots$$

i.e. starting from the first column, if the pairs are rearranged in a two-dimensional matrix form, the diagonal reading goes up and right.
(c) Define $f : \mathbb{N} \to \mathbb{Z}$ so that $f(1) = 0$ and $f(2i) = i$ and $f(2i+1) = -i$ for each $i \geq 1$. Function f is ONTO and 1-1. Given that the countable union of countable sets is countable so is $\mathbb{N}$, $\{0\}$ and $-\mathbb{N}$.
(d) There can be two possibilities for X: it can be countably infinite or finite. In the former case there is an ONTO and 1-1 function $f : \mathbb{N} \to X$ and we are done. For the latter case there is an $m \in N$ and then establish $g : \{1,2,\ldots,m\} \to X$. Then we define a map f $f : \mathbb{N} \to X$ and we are done. for $1 \leq i \leq n$, $f(i) = g(i)$; for other values of i, such that $i \geq n+1$, we pick an arbitrary $x \in X$ and have $f(i) = x$ for all $i \geq n+1$. The map defined is thus ONTO but not 1-1.
For the converse we assume that there is a onto map f such that $f : \mathbb{N} \to X$ and we are going to construct g an ONTO and 1-1 map (function) g such that $g : \{1,2,\ldots,m\} \to X$ or $g : \mathbb{N} \to X$. We start with $f(1)$ and set $g(1) = f(1)$. By (induction, yet to be defined) say we have generated $g(1), g(2), \ldots, g(n)$. We are thus about to generate $g(n+1)$. Construct $X_n = \{j \in \mathbb{N} : f(j) \neq g(i), i = 1,2,\ldots,n\}$. We thus collect in X_n, values j such that $f(j)$ is not equal to a $g(i)$ yet. We thus similary discard $f(j)$ values that have been assigned to $g(1),\ldots,g(n)$. X_n is either empty and thus the g map is 1-1 and ONTO and X is finite or X_n has a smallest elements k and thus $g(n+1) = f(k)$, and $f(k)$ is different from all the previous g values. This is either going to terminate with X being finite or we go through all the f values to obtain a 1-1 and ONTO map g for a countably infinite X.
(e) Let $A = \{X_i : i \in \mathbb{N}\}$ be a countable collection of countable sets X_i. Because the X_i are countable we have $g_i : \mathbb{N} \to X_i$. We further define $g : \mathbb{N} \times \mathbb{N} \to \cup_i X_i$, where $g(i,k) = g_i(k)$. Then there is a 1-1 and ONTO function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ with

$$g \circ f : \mathbb{N} \to \cup_i X_i.$$

(f) Let A be a countable collection of subsets of $\mathbb{N}$ i.e. $A \subset \mathbb{P}(\mathbb{N})$, with $A = \{A_n \subset \mathbb{N} : n \in \mathbb{N}\}$. Define then a subset B of $\mathbb{N}$. $B = \{n \in \mathbb{N} : n \notin A_n\}$. The $B \neq A_n$ for every n. This is because if there is a j in B then $j \notin A_j$; or $j \notin B$ and thus $j \in A_j$. Thus $B \notin A$ and thus no countable collection of subsets of $\mathbb{N}$ includes all subsets of $\mathbb{N}$ and the result follows. ■

Exercise 39.

(a) Show that the open interval of real numbers $(0,1)$ is uncountable. Use Cantor's diagonalization.
(b) A Lattice point $p=(x,y)$ is a point on the Cartesian plane $\mathbb{R}\times\mathbb{R}$ whose x and y coordinates are integer. You are given five points p_1,p_2,p_3,p_4,p_5. Each pair is connected with a line segment. Show that at least one of those line segments includes a lattice point (not necessarily one of the p_i).
(c) Out of 25 students at least three of them were born on the same month.
(d) Out of 899 students at least three of them share the same birthday.

Solution.
Let the reals be countably infinite (and thus countable). Write them down in the form

$$\begin{aligned} z_1 &= .z_{11}z_{12}z_{13}\dots \\ z_2 &= .z_{21}z_{22}z_{23}\dots \\ z_3 &= .z_{31}z_{32}z_{33}\dots \\ \dots &\quad \dots \end{aligned}$$

where the digits have double subscripts. Construct c as follows.

$$c = .c_1c_2c_3\dots$$

Real number c is such that $0<c<1$. if $z_{ii}\neq 1$, then $c_i=1$, if $z_{ii}==1$, then $c_i=3$. Thus c is not equal to any of the $z_1,z_2,\dots$. Therefore the reals in open interval cannot be listed, and they are uncountable. By extension $\mathbb{R}$ is uncountable.
The 5 points can have odd or even x and odd or even y. There are four possibilities. By the pigeonhole (Dirichlet) principle at least two points say p_1,p_2 have their x points the same parity and likewise for y's points. Consider the middle point of a segment. $(\frac{p_1(x)+p_2(x)}{2},\frac{p_1(y)+p_2(y)}{2})$. The same-axis coordinates have the same parity so their sum is even and one half of it integer. The middle point is thus a lattice point. ■

Exercise 40.

Pigeonhole (Dirichlet) principle. In a room there are 8 people. At least two of them were born on the same day of the week. Show the result.
In other words the pigeonhole principle is defined formally as follows. Let function f be defined $f : X \to Y$, with $|X| > |Y|$ then there are two elements $a, b \in X$ such that $a \neq b$ and $f(a) = f(b)$.

Solution.
Create 7 bins after the names of the week. Place 7 of the 8 in the appropriate bin. Either we have found one pair that was born on the same day of the week or each bin has one person only. The 8th person resolves the question. ■

Exercise 41.

Dirichlet (pigeonhole) principle. Let $m,n \in \mathbb{N}$. If $nm+1$ objects are split into n groups, then at least one of those group will contain at least $m+1$ objects.

Solution.
Let m_i be the number of object in group i, $i = 1,\ldots,n$. We have

$$\sum_i m_i = nm+1.$$

Let us assume that each group contains at most m objects. Then $m_i \leq m$. We then have.

$$\sum_i m_{i=1}^n = nm+1 \Leftrightarrow nm+1 = \sum_{i=1}^n m_i \Leftrightarrow nm+1 \leq \sum_{i=1}^n m \Leftrightarrow nm+1 \leq \sum_{i=1}^n m \Leftrightarrow nm+1 \leq nm,$$

which is false, since $nm+1 > nm$. ■

Exercise 42.

Show the following. Let $f : X \mapsto Y$ and let $|X| > |Y|$ for two finite sets X, Y. Then there are two elements $x_1, x_2 \in X$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2)$.
The proof is by induction on the size of X i.e. the cardinality of X. We state the result as a predicate $P(n)$.

> **Theorem 1.5**
>
> *For any finite set X and finite set Y such that $|X| > |Y|$ and $|X| = n$, if $f : X \mapsto Y$, then there are distinct elements of X, $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$.*

Solution.
Base case $n = 2$. If $|X| > |Y|$ and $|X| = 2$ then $|Y| = 1$ or $|Y| = 0$. Since there is an f then $|Y| = 1$ so that for $x \in X$ we have $f(x) \in Y$. Thus the two elements of X must map to the only element of Y.
Inductive hypothesis: Suppose the theorem is true for some arbitrary but fixed n.
Inductive step: We prove that if $|X| > |Y|$ and $|X| = n+1$, if $f : X \mapsto Y$, then there are distinct elements of X, $x_1 \neq x_2$ such that $f(x_1) = f(x_2)$.
Pick an arbitrary x_1 of X. There are two possibilities for x_1:
(a) there is $x_3 \in X$ such that $x_1 \neq x_3$ but $f(x_1) = f(x_3)$. In this case induction is complete; we are done. The pair $(x_1, x_2) = (x_1, x_3)$.
(b) there is NO $x_3 \in X$ such that $x_1 \neq x_3$ leads to $f(x_1) = f(x_3)$; in fact for every $x_3 \neq x_1$ we have $f(x_1) \neq f(x_3)$. This means that x_1 is the only element of X mapped to $f(x_1)$. Form them $X_1 = X - \{x_1\}$, and $Y_1 = Y - \{f(x_1)\}$. Apply the induction hypothesis on X_1, Y_1 and function $f_1 : X_1 \mapsto Y_1$, with $|X_1| = n-1$. Function f_1 is identical to f except that is not defined on x_1 that is missing from X_1. Then there exists $x_4, x_5 \in X_1$ such that $x_4 \neq x_5$ and yet $f_1(x_4) = f_1(x_5)$. But then $f(x_4) = f(x_5)$, and thus for function f we have found $(x_1, x_2) = (x_4, x_5)$ such that $f(x_1) = f(x_2)$ wince $f(x_1) = f_1(x_1) = f_1(x_4) = f_1(x_5) = f_1(x_2) = f(x_2)$. This completes the proof. ■

1.8 Numbers

Exercise 43.

Convert denary integer $n \geq 0$ into binary. (Right to left derivation.)

Solution.

```
   Convert2Binary (n)                  //Digits are generated right to left
Step 0.  N=n; if (n<2) return(n);      //from least to most significant
STEP 1. while (N > 0)  {
STEP 2. If ( a = N % 2 )  // a is set to 1 for odd N; to 0 for even N
          out(a) ;       // print bit
STEP 3. N = (N - a) / 2;  // N-a is always even; Division by 2 is shift right
STEP 4. if (N==0) exit;
        }
```

We provide an example for $n = 112$.

```
   Convert2Binary (n=112)                Output
      N= n = 112;
      true(N > 0) ;
      a= 112%2   = 0 ; out(a);                    0
      N= (N-a)/2 = 56;
      true(N > 0) ;
      a=  56%2   = 0 ; out(a);                   00
      N= (N-a)/2 = 28;
      true(N > 0) ;
      a=  28%2   = 0 ; out(a);                  000
      N= (N-a)/2 = 14;
      true(N > 0) ;
      a=  14%2   = 0 ; out(a);                 0000
      N= (N-a)/2 =  7;
      true(N > 0) ;
      a=   7%2   = 1 ; out(a);                10000
      N= (N-a)/2 =  3;
      true(N > 0) ;
      a=   3%2   = 1 ; out(a);               110000
      N= (N-a)/2 =  1;
      true(N > 0) ;
      a=   1%2   = 1 ; out(a);              1110000
      N= (N-a)/2 =  0;
      false(N > 0) ;              bin(112) = 1110000
```

■

Exercise 44.

Convert denary integer $n \geq 0$ into binary. (Left to right derivation.)

Solution.

```
   Convert2Binary (n)  //Digits are generated left to right
Step 0.  N = n ; if (n<2) return(n);
Step 1.  t = 1 ; m = 1;
Step 2.  while ( t <= N ) {
Step 3.     m = t;
Step 4.     t = 2 * t;
         }
Step 5.
STEP 6.  while (m > 0)  {
STEP 7.    If ( N >= m ) {
STEP 8.       out(1) ;       // print  1 ; m fits into N
STEP 9.       N = N - m;
STEP 10.      m = m / 2;
STEP 11.   }
STEP 12.   else {
STEP 13.      out(0);        // print  0 ; m does not fit into N
STEP 14.      m = m / 2
STEP 15.   }
STEP 16. }
STEP 17.  If (N==0) exit;
```

We provide an example for $n = 112$.

```
   Convert2Binary (n=112)                Output
      N= n = 112;
      t = 1 ; m = 1;  // Find largest power of 2 <= N
      true( t <= N )  // and set it to m
      m = 1 ;
      t= 2*t = 2;
      true( t <= N)
      m = 2 ;
      t= 2*t = 4;
      true( t <= N)
      m = 4 ;
      t= 2*t = 8;
      true( t <= N)
      m = 8 ;
      t= 2*t = 16;
      true( t <= N)
      m = 16;
      t= 2*t = 32;
      true( t <= N)
      m = 32;
      t= 2*t = 64;
      true( t <= N)
      m = 64;
      t= 2*t = 128;
      false( t<= N);  // Largest power of 2 <= 112 is  m=64
      //                                              |
      // N= 112 ; m =64 -------------------------------|
      true( m > 0)
        true( N >= m)
          out(1);                        1
          N = N-m=112-64 = 48;
          m = m/2=32;
      true( m > 0)
        true( N >= m)
          out(1);                        11
```

```
    N = N-m=48-32  = 16;
    m = m/2=16;
true( m > 0)
  true( N >= m)
    out(1);                        111
    N = N-m=16-16  =  0;
    m = m/2= 8;
true( m > 0)
  false( N >= m)
    out(0);                        1110
    m = m/2= 4;
true( m > 0)
  false( N >= m)
    out(0);                        11100
    m = m/2= 2;
true( m > 0)
  false( N >= m)
    out(0);                        111000
    m = m/2= 1;
true( m > 0)
  false( N >= m)
    out(0);                        1110000
    m = m/2= 0;
false( m > 0)
    exit;              bin(112) = 1110000
```

■

Chapter 2

Boolean algebra and proofs

2.1 Short review on boolean algebra and functions

2.1.1 Boolean algebra

Boolean functions. A boolean function f on n parameters (variables) has domain $\{0,1\}^n$ and codomain (range) $\{0,1\}$.
Moreover we associate 1 with T for True and 0 with F for False.
Fundamental boolean algebra functions. The three fundamental boolean functions are NOT(x), OR(x,y), and AND(x,y) for negation, disjunction and conjunction respectively. They are defined on one or two parameters. We refer to the fundamental boolean algebra functions as logical operations. A logical operation is denoted by a symbol or short name known as an operator.
Funamental boolean algebra functions: logical operators. The list of fundamental boolean algebra functions is given below. Each function maps to a symbol defined as an operator. The operation implied by the operator (symbol) is the boolean function. An operator with one argument (parameter) is a unary operator; with two it is a binary operator and we describe x as its left operand (argument) and y its right operand (argument). The symbol operatior is described as an infix operator appearing between its two arguments (or null first argument and second argument) for negation. We also display the corresponding prefix operator where the operator precedes its one or two arguments.

Name of function	Prefix operator	Symbol Operator
Negation	NOT(x)	$(\neg x)$
Disjunction (inclusive)	OR(x,y)	$(x \vee y)$
Conjunction	AND(x,y)	$(x \wedge y)$
Disjunction (exclusive)	XOR(x,y)	$(x \oplus y)$

2.1.2 Additional boolean functions

Exclusive disjunction. The exclusive disjunction function denoted by XOR(x,y) or $x \oplus y$, is not a fundamental boolean algebra function. It can be expressed in term of the three fundamental functions. Then

$$x \oplus y = (x \vee y) \wedge \neg(x \wedge y)$$

which is equal to (or equivalent to)

$$x \oplus y = (x \wedge \neg y) \vee (y \wedge \neg x).$$

Conditional. It is also known as implication.

$$\text{Implication (or conditional)} \quad x \Rightarrow y$$

Biconditional. The $x \Leftrightarrow y$ is a biconditional and is defined as $x \Rightarrow y \wedge y \Rightarrow x$.

Equivalence (or biconditional) $x \Leftrightarrow y$

2.1.3 Logical equivalence of functions.

Two functions are logically equivalent if their trueness values are the same that is, $f = g$. The symbol $\equiv$ stands for "logically equivalence" or "logically equivalent" and can be used in place of $=$.

x	y	$\neg x$	$x \vee y$	$x \wedge y$	$x \oplus y$	$x \Rightarrow y$	$x \Leftrightarrow y$
0	0	1	0	0	0	1	1
0	1	1	1	0	1	1	0
1	0	0	1	0	1	0	0
1	1	0	1	1	0	1	1

2.1.4 Laws of boolean functions

In the formulae below we use x, y, z for boolean variables.

	Disjunction	Conjunction
Associative	$(x \vee y) \vee z = x \vee (y \vee z)$,	$(x \wedge y) \wedge z = x \wedge (y \wedge z)$
Commutative	$x \vee y = y \vee x$,	$x \wedge y = y \wedge x$
Distributive	$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$	$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$
Identity	$x \vee F = x$	$x \wedge T = x$
Inverse	$x \vee (\neg x) = T$	$x \wedge (\neg x) = F$
Idempotent	$x \vee x = x$	$x \wedge x = x$
Domination	$x \vee T = T$	$x \wedge F = F$
Double Negation	$\neg(\neg x) = x$	$\neg(\neg x) = x$
De Morgan	$\neg(x \vee y) = (\neg x) \wedge (\neg y)$,	$\neg(x \wedge y) = (\neg x) \vee (\neg y)$
Absorptive	$x \vee (x \wedge y) = x$,	$x \wedge (x \vee y) = x$.

Example 2.1.1

Consider $P(n) : f(n) = \frac{6n+10}{3n+2}$ for integer n and $Q(n) : f(n) \neq 2$ for integer n. We are asked to prove $P(n) \Rightarrow Q(n)$. Note that $\neg Q(n) : f(n) = 2$.

Solution. Starting with $P(n)$ and $\neg Q(n)$ we have

$$f(n) = \frac{6n+10}{3n+2} = 2 \Leftrightarrow 6n+10 = 6n+4 \Leftrightarrow 10 = 4 \Leftrightarrow 6 = 0.$$

The contradiction C can be $10 = 4$ or $6 = 0$ and has nothing to do with $P(n)$ or $Q(n)$. ■

2.1.5 Propositions

There are simple and compound propositions.

Definition 2.1 **Proposition**

A proposition (also known as simple proposition) is a statement that is either true or false.

An implication of this definition is that a proposition can't be both true and false. Several times the terms proposition and statement are used interchangeably.

The statement of a proposition can depend on the values of variables.

Definition 2.2 **Predicate or predicate proposition**

A **predicate** is a proposition whose truth value is dependent on the value of one or more variables.

The statement of a proposition can depend on the values of variables and it is known as a predicate or predicate proposition. When the values are quantified (assigned values), then the predicate becomes a proposition.

Definition 2.3 **Compound proposition**

A compound proposition is formed by joining two or more propositions with logical operators such as $\wedge, \vee, \neg, \oplus, \Rightarrow, \Leftrightarrow$.

Definition 2.4 **Universe of discourse**

For a predicate proposition the set of values of its variable is the universe of discourse (UD).

For a predicate proposition $P(x)$ that maps to a boolean variable, the UD of $\{0,1\}$. For a predicate proposition $P(x,y)$ that maps to a boolean function the UD of $\{(0,0),(0,1),(1,0),(1,1)\}$. For a predicate proposition $P(n)$ that depends on a natural number, $\mathbf{UD} = \mathbb{N}$.

2.1.6 Compound propositions and implications

The components of an implication or conditional: Antecedent and consequent.
The implication or conditional

$$P \Rightarrow Q$$

P is the antecedent (hypothesis), and Q is the consequent (conclusion). The implication $P \Rightarrow Q$ of say two propositions P and Q used as the antecedent and consequent is also a (compound) proposition.
Reading an implication. We can read the implication as defined above in four different ways: "If P then Q", or "P implies Q", or "P only if Q", or "Q if P". In an implication P is sufficient for Q. If P occurs i.e. it is true then the (true) implication implies that Q will be true. In an implication Q is necessary for P that is, Q is necessary for P to be true in the sense that the trueness of P guarantees Q to be true.
Trueness of an implication. The implication $\mathbf{P \Rightarrow Q}$ is false if P is true and Q is false, and it is true in all other cases. The implication $\mathbf{P \Rightarrow Q}$ is equivalent to $\neg P \vee Q$. Thus if P is F, then Q is T: we then can say the implication is vacuously true.

Definition 2.5 **Converse of an implication**

The converse of implication $P \Rightarrow Q$ is another implication: $Q \Rightarrow P$.

An implication is a proposition. The converse of an implication is also a proposition.

Definition 2.6 **Inverse of an implication**

The inverse of implication $P \Rightarrow Q$ is another implication: $\neg P \Rightarrow \neg Q$.

An implication is a proposition. The inverse of an implication is also a proposition.

Definition 2.7 **Contrapositive of an implication**

The contrapositive of implication $P \Rightarrow Q$ is another implication: $\neg Q \Rightarrow \neg P$.

An implication is a proposition. The contrapositive of an implication is also a proposition.

Definition 2.8 **Contradiction**

Contradiction is a compound proposition that is always F. A contradiction on x is defined as $x \wedge (\neg x)$.

Definition 2.9 **Tautology**

Tautology is a compound proposition that is always T. A tautology on x is defined as $x \vee (\neg x)$.

Reading a biconditional. The proposition

$$P \Leftrightarrow Q$$

is a biconditional (proposition) statement. It reads P only if Q and P if Q. Or equivalently If P, then Q, and If Q, then P.

Equivalence of implications. In the formulae below we use p, q, r for propositions. A proposition can substitute for a variable but then equality becomes equivalence.

Implication	$p \Rightarrow q \equiv \neg p \vee q$	
Contrapositive	$p \Rightarrow q \equiv \neg q \Rightarrow \neg p$	Equivalence contrapositive and implication
Reduction Ad Absurdum (Contradiction)	$p \Rightarrow q \equiv (p \wedge (\neg q) \Rightarrow F.$	

2.1.7 Conjunctive and disjunctive normal forms

Definition 2.10 **Conjunctive normal form**

Let x_i be variables; x_i and $\neg x_i$ are then referred to as literals. Then an expression $E_1 \vee \ldots \vee E_k$ where E_i is literal is called a clause C. A boolean function $f : \{0,1\}^n \to \{0,1\}$, is in conjunctive normal form (CNF) if it is of the form

$$f(x_1, \ldots, x_n) = C_1 \wedge \ldots \wedge C_m,$$

where C_i is a clause. Every boolean function can be expressed in CNF.

One can similarly define a disjunctive normal form (DNF).

Definition 2.11 **Disjunctive normal form**

Let x_i be variables; x_i and $\neg x_i$ are then referred to as literals. Then an expression $E_1 \wedge \ldots \wedge E_k$ where E_i is literal is called a clause D. A boolean function $f : \{0,1\}^n \to \{0,1\}$, is in disjunctive normal form (DNF) if it is of the form

$$f(x_1, \ldots, x_n) = D_1 \vee \ldots \vee D_m,$$

where D_i is a clause. Every boolean function can be expressed in DNF.

2.1.8 Boolean matrices

Definition 2.12 **Matrix**

A matrix A is a geometric object consisting of r rows and c columns. Then, $r \times c$ boolean values are arranged in the rows and columns of the matrix. Matrix A is said to have geometry $r \times c$ if its number of rows is r and its number of columns c. One may then use the notation

$$A = A_{r \times c} = (a_{ij}).$$

The element at row index i and column index j of A is denoted by a_{ij}.

Definition 2.13 **Boolean matrix**

A boolean matrix B is a matrix B where b_{ij} is either 0 or 1 only.

For two boolean matrices A, B one can define $A \wedge B$ and $A \vee B$ as element-wise operations. Likewise $C = A \times B$ is the boolean product $c_{ij} = \vee_k a_{ik} \wedge b_{kj}$. The geometry of A and B for C to be defined draw from ordinary matrix multiplication: for an A that is $n \times p$ and B that is $p \times m$, then C is defined and is $n \times m$. Properties of matrices translate to boolean matrices.

2.1.9 Quantifiers

Definition 2.14 **Existential quantifier**

Let $P(x)$ be a predicate proposition. The $(\exists x)P(x)$ is a proposition that is true when at least one element e of the Universe of Discourse satisfies $P(e)$ is true.

Definition 2.15 **Universal quantifier**

Let $P(x)$ be a predicate proposition. The $(\forall x)P(x)$ is a proposition that is true when every element e of the Universe of Discourse satisfies $P(e)$ is true.

Definition 2.16 **Argument vs Parameter**

In $P(x)$ of a predicate opposition but also of a function, x is the parameter. When we substite e for x as in $P(e)$, e becomes the argument of P.

Example 2.1.2

Let us assume that the universe of discourse for x, y below is the set of natural numbers. Is

$$(\forall x)(\exists y)P(x,y) \equiv (\exists y)(\forall x)P(x,y)?$$

Solution. Consider the function $s(x) = x+1$, the successor of x. Let $P(x,y) : y = x+1$ in other words $y = s(x)$. The first proposition states for all x there exists a successor of x denoted by y. The second proposition reads that tha there is a y that is the successor of all integers. Naturally the former is true, but the latter is false. In fact the former is one of Peano's axioms of natural numbers! ■

The best way to understand this is to think of a person and a name. Every person has a name. But for a given name it is not the case that all persons (of the UD) are given that give name.

The following definition of a limit serves also as an example of the use of quantifiers.

Definition 2.17 **Limit to zero for $n \to \infty$**

Let x_n be a sequence of real numbers, $n = 0, 1, \ldots, \infty$. The limit of x_n is 0 and we write $\lim_{n\to\infty} x_n = 0$ if for every $\varepsilon > 0$ there exists a (natural integer) number $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $|x_n| \leq \varepsilon$. In other words we have the following.

$$(\forall \varepsilon > 0)(\varepsilon \in \mathbb{R})(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(|x_n| < \varepsilon).$$

2.1.10 De Morgan's Laws for Predicates

Definition 2.18 **De Morgan's Law for $\forall$**

$\neg((\forall x)P(x)) \equiv ((\exists x)(\neg P(x)))$.

Definition 2.19 **De Morgan's Law for $\exists$**

$\neg((\exists x)P(x)) \equiv ((\forall x)(\neg P(x)))$.

To disprove a $(\forall x)P(x)$ it suffices to find a so that $P(a)$ is F.

2.1.11 Proof techniques

Method 2.1 **By case analysis**

For a proposition $(\forall x)P(x)$, we need to show that $P(x)$ is true. We split the universe of discourse into non-disjoint subsets and prove $P(x)$ is true by Method 4 or 5 for each (sub)set.

Method 2.2 **By universality**

For a proposition $(\forall x)P(x)$, we need show that $P(x)$ is true without any restriction on x.

Method 2.3 **By direct deduction**

For a proposition $(\forall x)(P(x) \Rightarrow Q(x))$, we start with a generic element of the universe of discourse and by assuming $P(x)$ is true we conclude $Q(x)$ is true.

Method 2.4 **By existence**

For a proposition $(\exists x)P(x)$, we need to find an element e of the universe of discourse with $P(e)$ be true. If however $(\exists\ !x)P(x)$, we need to find an x and prove that it is unique, in other words

$$(\exists x)(P(x) \wedge ((\forall y)(P(y) \Rightarrow (x = y)))).$$

In other words we first find one e such that $P(e)$ is true, and then establish that if there is a y such that $P(y)$ is true, the $y = e$. Alternatively we can establish that if y, z exist $P(y)$ and $P(z)$ are both true then $y = z$.

Method 2.5 **Disproving by a counterexample**

For a proposition $(\forall x)P(x)$, we need to show that $P(x)$ is false. We find e such that $P(e)$ is false. If we can't find such as e it means $P(x)$ is true for all elements x of the universe of discourse!

Method 2.6

By contrapositive] For a proposition $P \to Q$ the contrapositive is $\neg Q \to \neg P$. Thus instead of proving the more complicated $P \to Q$ to be true, we prove $\neg Q \to \neg P$, the less complicated, to be true!

Method 2.7 **Modus Ponens: Confirming the antecedent**

For a proposition that an implication $P \Rightarrow Q$ is, P is the antecedent, and Q is the consequent. If we confirm the antecedent, i.e. if P is true, then the consequent Q is true.

Method 2.8 **Modus Tollens: Denying the consequent**

For a proposition that an implication $P \Rightarrow Q$ is, P is the antecedent, and Q is the consequent. If we deny the consequent, i.e. if Q is false, then the antecedent P is false, since otherwise a true P would result into a true Q.

Method 2.9 **Reduction Ad Absurdum : by contradiction**

For a proposition $P \rightarrow Q$, we prove the proposition true by proving an equivalent proposition true. The equivalent proposition is

$$P \wedge (\neg Q) \Rightarrow F$$

That is starting with P and $\neq Q$ we conclude a False (F) statement (proposition). A more restrictive form is the following one.

$$\neg P \Rightarrow (Q \wedge (\neg Q))$$

Method 2.10 **Biconditionals**

For a proposition $P \Leftrightarrow Q$, prove separately $P \Rightarrow Q$ and $Q \Rightarrow P$ by using the previous methods.

Method 2.11 **Induction**

For a predicate propostion $P(n)$ in order to show $(\forall n \in \mathbb{N})P(n)$, show $P(0)$ and $(\forall n \in \mathbb{N})(P(n) \Rightarrow P(n+1))$.

2.1.12 Peano's axioms of Arithmetic

Definition 2.20 **Axioms of Arithmetic: Peano's axioms**

Peano's axioms define natural numbers. The set of natural numbers is denoted by N or $\mathbb{N}$.

$$\mathbb{N} = \{0, 1, 2, \ldots\}$$

2.1.13 Simplified Peano exposition

Axiom 2.1 **First Peano Axiom**

Zero (0) is a natural number.

Axiom 2.2 **Second Peano Axiom**

If n is a natural number, then its successor $s(n)$ is also a natural number.
(We prefer to write $n+1$ for the successor $s(n)$ of n.)

Therefore the set of natural numbers is closed under the successor function $s(n)$.

Axiom 2.3 **Third Peano Axiom**

Natural number zero is not the successor of a natural number.

Axiom 2.4 **Fourth Peano Axiom**

If n,m are natural numbers then $s(n) = s(m)$ if and only if $n = m$.

Function $s(n)$ is thus an injection (injective function). Therefore $n \neq m$ implies $s(n) \neq s(m)$. By contrapositive, it is then the case that $s(n) = s(m)$ implies $n = m$ since otherwise, if $n \neq m$, then $s(n) \neq s(m)$.

Axiom 2.5 **Fifth Peano Axiom: Induction**

If for a subset $A \subseteq \mathbb{N}$ of the natural numbers we have that (a) $0 \in A$, and (b) $\forall a \in X \Rightarrow s(a) \in A$, then $A = \mathbb{N}$.

The fifth axiom is usually understood or used as follows. Let $P(n)$ be a predicate defined on natural numbers n. If (a) $P(0)$ is true, and (b) for all natural numbers n we have $P(n) \Rightarrow P(s(n))$, then $P(n)$ is true for all natural numbers n.

2.1.14 Extensive Peano exposition

A theorem in mathematics is true because it is based and derived from simple axioms that are accepted as true or are usually true.

Theorem 2.1 **Well Ordered Set Principle**

Every non-empty subset of $\mathbb{N}$ *has a minimal element.*

There is a conjecture that is known as Goldbach's conjecture. It can be stated as follows.

Proposition 2.1 **Goldbach's Conjecture**

Every even integer greater than two is the sum of two prime numbers.

We can restate it in a more formal form as follows.

Proposition 2.2 **Goldbach's Conjecture**

$$\forall n \in \mathbb{N}. \exists p \in \mathbb{N}. \exists q \in \min\{N\} : 2/n \wedge Prime(p) \wedge Prime(q) \wedge n = p + q.$$

According to Wikipedia, Proposition 2.2 is true for all integers up to about $4 \cdot 10^{18}$; no one knows whether Proposition 2.2 is indeed true as stated. Every proposition is true or false with reference to a space we first define axiomatically and then build on by establishing more theorems. Proposition 2.3 below is defined in the form of predicate $P(n)$.

Proposition 2.3

P(n): For natural number n, integer $n^2 + n + 5$ is prime.

Another way to say this is as follows.

Remark 2.1

P(n): $\forall n \in \mathbb{N}$, $n^2 + n + 5$ is prime.

Remark 2.2

Reminder: A prime number is a natural number greater than 0 whose is divisible by one and itself. (Prime numbers that are integers have divisors the two units $+1$ and -1, and the products of the units and the number itself.)

Remark 2.3

P(n) is a predicate that depends on n. It is not a function. P(n) is either true or false depending on n.
Moreoever P(5) reads
For natural number 5, *integer* 5^2+5+5 *is prime.*
Obviously $P(5)$ is false.
Moreoever P(3) reads
For natural number 3, *integer* 3^2+3+5 *is prime.*
Obviously $P(3)$ is true.

2.1.15 Ordinary induction

Ordinary (weak) induction. It can be stated in the form of the theorem below. The proof of Theorem 2.2 follows in the problem section. An alternative formulation for Theorem 2.2 can be given that excludes zero from the set of natural numbers. This set is then denoted by $\mathbb{N}^*$. This is Theorem 2.3 below. If Theorem 2.2 is TRUE it also follows that Theorem 2.3 is also TRUE.

Theorem 2.2 **Ordinary Induction**

Let $A \subseteq \mathbb{N}$. *Let*

- *(i)* $0 \in A$, *and*
- *(ii) whenever* $k \in A$ *then* $k+1 \in A$.

The two conditions (i) and (ii) imply $A = \mathbb{N}$.

Theorem 2.3 **Alternative Ordinary Induction**

Let $A \subseteq \mathbb{N}^*$.

- *(i)* $1 \in A$, *and*
- *(ii) whenever* $k \in A$ *then* $k+1 \in A$. *The two conditions (i) and (ii) imply* $A = \mathbb{N}^*$.

2.1.16 Strong induction

Strong induction derives from and is equivalent to weak (ordinary) induction. It is called strong because the "whenever $k \in A$ then $k+1 \in A$" is replaced by a richer "whenever $\{0,1,\ldots,k\} \subseteq A$ then $k+1 \in A$.

Theorem 2.4 **Strong Induction**

Let $A \subseteq \mathbb{N}$. *Let*

- *(i)* $0 \in A$, *and*
- *(ii) whenever* $\{0,1,\ldots,k\} \subseteq A$ *then* $k+1 \in A$.

The two conditions (i) and (ii) imply $A = \mathbb{N}$.

The alternative form of strong induction follows.

Theorem 2.5 **Alternative strong Induction**

Let $A \subseteq \mathbb{N}^*$. Let

- *(i) $1 \in A$, and*
- *(ii) whenever $\{1, 2, \ldots, k\} \subseteq A$ then $k+1 \in A$.*

The two conditions (i) and (ii) imply $A = \mathbb{N}^*$.

2.1.17 Strong versus weak induction

In computing we prove the validity of proposition P that depend on an integer (natural number) variable denoted by n. In other words we are interested in proving

$$\forall n.P(n),$$

implying equivalently that

$$\forall n \in \mathbb{N}.P(n).$$

Thus with induction, we consider set A of Theorem 2.2 or Theorem 2.4 to be the subset of $\mathbb{N}$ for which $P(n)$ is true

$$\forall n \in A.P(n),$$

or equivalently

$$A = \{n \in \mathbb{N} : P(n)\},$$

and then proving that $A = \mathbb{N}$.

In this discussion we assume that the base case is 0. Besides the alternative formulation where base case is 1, we can extend it so that the base case is any n_0 such that $n_0 > 0$ and $n_0 \in \mathbb{N}$. The two methods are equivalent. We formulated them using $P(n)$ instead of set A (that is dependent of $P(n)$. We thus obtain the following.

Theorem 2.6 **Ordinary induction method**

Let $P(n)$ be a proposition dependent on natural number n, $n \in \mathbb{N}$. Let (i) $P(0)$ *be true, and* (ii) $\forall n,\ P(n) \Rightarrow P(n+1)$.
The two conditions (i) and (ii) imply $\forall n.P(n)$, that is $P(n)$ is true for all $n \in \mathbb{N}$.

Theorem 2.7 **Strong induction method**

Let $P(n)$ be a proposition dependent on natural number n, $n \in \mathbb{N}$. Let
(i) $P(0)$ *be true, and*
(ii) $\forall n,\ P(0) \wedge P(1) \wedge \ldots \wedge P(n) \Rightarrow P(n+1)$.
The two conditions (i) and (ii) imply $\forall n.P(n)$, that is $P(n)$ is true for all $n \in \mathbb{N}$.

We can restate both induction methods with $P(n)$ as the consequent instead of $P(n+1)$. Moreover we rewrite the strong induction case (ii) in a way that eliminates the dots.

Theorem 2.8 **Ordinary induction method restated**

Let $P(n)$ be a proposition dependent on natural number n, $n \in \mathbb{N}$. Let
(i) $P(0)$ *be true, and*
(ii) $\forall n.n \geq 1.\ P(n-1) \Rightarrow P(n)$.
The two conditions (i) and (ii) imply $\forall n.P(n)$, that is $P(n)$ is true for all $n \in \mathbb{N}$.

Theorem 2.9 **Strong induction method restated**

Let $P(n)$ be a proposition dependent on natural number n, $n \in \mathbb{N}$. Let
(i) $P(0)$ *be true, and*
(ii) $\forall n,\ ((\forall k \in \mathbb{N}\,(k < n \Rightarrow P(k))) \Rightarrow P(n))$.
The two conditions (i) and (ii) imply $\forall n.P(n)$, that is $P(n)$ is true for all $n \in \mathbb{N}$.

The two forms of induction are equivalent. Consider $Q(n)$ defined as follows Then the part (ii) of the strong induction method for $P(n)$ is equivalent to part (ii) of the ordinary induction method for $Q(n)$, and also noting that $Q(0) = P(0)$ and thus part (i) does not change for either $P(n)$ or $Q(n)$!

$$Q(n) = P(0) \wedge P(1) \wedge \ldots \wedge P(n)$$

$$\begin{array}{lrcl} \forall n, & P(0) \wedge P(1) \wedge \ldots \wedge P(n) & \Rightarrow & P(n+1) \Leftrightarrow \\ & \forall n Q(n) & \Rightarrow & P(n+1) \Leftrightarrow \\ & \forall n Q(n) & \Rightarrow & P(0) \wedge P(1) \wedge \ldots \wedge P(n) \wedge P(n+1) \Leftrightarrow \\ & \forall n Q(n) & \Rightarrow & Q(n+1) \end{array}$$

2.1.18 Different base case

Theorem 2.10 **Ordinary Induction : another base case**

Let $A \subseteq \mathbb{N}$. Let $a \in \mathbb{N}$ be a natural number with $a > 1$. Let

- *(i) $a \in A$, and*
- *(ii) $\forall k \in \mathbb{N}. k \geq a (k \in A \Rightarrow k+1 \in A)$.*

The two conditions (i) and (ii) imply the following.

$$A = \{n \in \mathbb{N} : n \geq a\}.$$

Theorem 2.11 **Strong Induction : another base case**

Let $A \subseteq \mathbb{N}$. Let $a \in \mathbb{N}$ be a natural number with $a > 1$. Let

- *(i) $a \in A$, and*
- *(ii) whenever $\{0, 1, \ldots, k\} \subseteq A$ then $k+1 \in A$, which can also be stated more formally as*
- *(ii') $\forall k \in \mathbb{N}. k \geq a (\{a, \ldots, k\} \subseteq A \Rightarrow k+1 \in A)$.*

The two conditions (i) and (ii) imply the following.

$$A = \{n \in \mathbb{N} : n \geq a\}.$$

The alternative form of strong induction follows.
In the remainder, the word induction can refer to either of them.

2.1.19 Why does induction work?

Another way to explain induction is the following discussion. On the left hand side we have what is known to be true. Part (i) of weak induction is $P(0)$. Part (ii) of weak induction are all the following expressions under column Given. What one derives is the right hand side under Derived. Thus $P(0)$ by way of is trueness that is given also appears under Derived obviously. Any proposition Derived in a prior row can be used for a derivation in a later row. Thus $P(0$ along with $P(0) \Rightarrow P(1)$ derives $P(1)$. The conclusion summarizes the right column's derivations i.e. $P(0), \ldots, P(n), \ldots$.

Given i.e. true	Derived/Obtained to be true
$P(0)$	$P(0)$
$P(0) \Rightarrow P(1)$,	$P(1)$
$P(1) \Rightarrow P(2)$,	$P(2)$
$P(2) \Rightarrow P(3)$,	$P(3)$
$\ldots$	$\ldots$
$P(n-1) \Rightarrow P(n)$,	$P(n)$
$P(n) \Rightarrow P(n+1)$.	$P(n+1)$
Conclusion:	$\forall n.P(n)$

Likewise for strong induction we have the following

Given i.e. true	Derived/Obtained to be true
$P(0)$	$P(0)$
$P(0) \Rightarrow P(1)$,	$P(1)$
$P(0) \wedge P(1) \Rightarrow P(2)$,	$P(2)$
$P(0) \wedge P(1) \wedge P(2) \Rightarrow P(3)$,	$P(3)$
$\ldots$	$\ldots$
$P(0) \wedge \ldots \wedge P(n-1) \Rightarrow P(n)$,	$P(n)$
$P(0) \wedge \ldots \wedge P(n)) \Rightarrow P(n+1)$,	$P(n+1)$
Conclusion:	$\forall n.P(n)$

2.2 Propositional calculus

Exercise 45.

$$P : 2+2=4. \tag{2.1}$$
$$Q : 2+2=3. \tag{2.2}$$
$$R : P \vee Q. \tag{2.3}$$
$$S : (2+2=4) OR (2+2=3). \tag{2.4}$$

Classify each one of the four propositions P, Q, R and S as True or False. You may assume that all integers are denary (i.e. radix 10).

Solution.
Obviously P is T and Q is F.
$P \vee Q$ that is proposition R is T as well.
Finally S is True as well; it is another way to express R. From now on, we will be using T and F for True and False. ■

Exercise 46.
The following C/C++ fragment of a program is given.

```
1.  i = 2;
2.  j = 3;
3.  total = 10;
4.  if (i > 1) && (j < 4)
5.    total = total + 2*i +j;
6.  return(total)
```

(a) Does the conditional of line 4 evaluate to T or F i.e. true or false?
(b) Is line 5 going to be executed ? Explain.
(c) What is the value returned by total? Explain.

Solution.
(a) Let $A = (i > 1)$ and $B = (j < 4)$. Then $C = A \wedge B$ is a form to express the condition of line 4. A is T since $i = 2 > 1$. Moreover B is T as well since $j = 3 < 4$. Thus the conjunction C is T as well.
(b) Line 5 will then be executed, and the value of total will be updated by increasing it by $2i + j = 2 \cdot 2 + 3 = 12$, and thus the new value would be $10 + 12 = 22$.
(c) The value of total returned would be 22. ■

Exercise 47.
Let $a = 10$, $b = 20$ and $c = 30$. Evaluate the expression $!(a > b)\&\&(b > a)$.
Furthermore evaluate the expression $!((a < b)\&\&(c \leq (a+b)))$.

Solution.
The first expression first. For $a = 10, b = 20$, $(a > b)$ is F thus $!(a > b)$ is T. $b > a$ is T. The compound expression is thus T.
$(a < b)$ is T. $c \leq (a+b)$ is true since $c = a+b = 30$. The conjunction $((a < b)\&\&(c \leq (a+b))$ is T and the expression is F. ■

Exercise 48.

(a) A conditional reads: "if it snows, then Alex will get sick". We are told "It did not snow". Discuss the validity of "Alex will get sick". Is the implication T or F then?

(b) The conditional reads: "if it snows, then Alex will get sick". We are told "Alex did not get sick". Discuss the validity of "It did not snow". Is the implication T or F then?

(c) Find the negation of $P \Rightarrow Q$.
(d) "If cats are bats, then bats have four legs." What is the negation of this conditional (implication)?
(e) What is the converse of implication $P \Rightarrow Q$?
(f) What is the inverse of implication $P \Rightarrow Q$?
(g) What is the contrapositive of implication $P \Rightarrow Q$?

Solution.
(a) Even if it did not snow, Alex might or might not get sick. If he gets sick it won't be because it snowed. The antecedent is F, the consequent can be F or T. The implication is T, as there is nothing that evaluates it as F.

(b) The statement (alternative name for a proposition) "It did not snow" is T. Had it snowed, Alex would have gotten sick but we have been told that Alex did not get sick. Therefore proposition "It did not snow" is T. Moreover the implication is T.

(c) First $P \Rightarrow Q$ is (logically) equivalent to $\neg P \vee Q$. Thus its negation is the negation of the disjunction i.e. $\neg(\neg P \vee Q)$ which is $P \wedge \neg Q$.

(d) Let P be "cats are bats" and let Q be "bats have four legs". The implication is then $P \Rightarrow Q$. Its negation is $P \wedge \neg Q$ i.e. "cats are bats and bats DONOT have four legs".
(e) The converse of implication $P \Rightarrow Q$ is another implication: $Q \Rightarrow P$. An implication is a proposition. The converse of an implication is also a proposition.
(f) The inverse of implication $P \Rightarrow Q$ is another implication: $\neg P \Rightarrow \neg Q$. An implication is a proposition. The inverse of an implication is also a proposition.
(g) The contrapositive of implication $P \Rightarrow Q$ is another implication: $\neg Q \Rightarrow \neg P$. An implication is a proposition. The contrapositive of an implication is also a proposition. ■

Exercise 49.
(a) Let $Z = a > b$. Consider $\neg Z$. Write it down. We are told $a = 5$ and $b = 10$. Evaluate Z and $\neq Z$.
(b) Let P : Triangle ABC has two sides equal to each other.
Let Q : Triangle ABC is equilateral.
Let R : Triangle ABC is isosceles.
Examine the trueness of the following
(b1) $P \Rightarrow Q$.
(b2) $P \Rightarrow R$.
(b3) $Q \Rightarrow R$.
(b4) $R \Rightarrow Q$.
(b5) $R \Rightarrow P$.
(b6) $R \Leftrightarrow P$.

Solution.
(a) If $a = 5, b = 10$ it means $a \leq b$. This means $Z = a > b$ is false i.e. Z evaluates to F. Likewise $\neg Z$ evaluates to T.
Implication (b1) might or might not be true since we have no information about the third side of the triangle.
Implication (b2) is true since an isosceles triangle has two sides equal to each other.
Implication (b3) is true since an equialateral triangle has all of its sides equal to each other and thus it is an isosceles triangle on any two of its sides.
Implication (b4) might or might not be true since we have no information about the third side of the isosceles triangle.
Implication (b5) is true since an isosceles triangle has two sides equal to each other.
Biconditional (b6) is true by way of (b5) and (b2). ■

Exercise 50.
Define ADD(x,y) $= 0$ if $x = y = 0$ or $x = y = 1$ and 1 otherwise. Define MUL(x,y) $= 0$ if $x = y = 1$ and 0 otherwise. Let $x = x_1x_0$ and $y = y_1y_0$ be two 2-digits binary numbers, representing x, y. Let $z = x + y = z_2z_1z_0$. Show the relationship between z_i and x_i and y_i.

Solution.
Naturally $z_0 = ADD(x_0, y_0) = x_0 \oplus y_0$. Furthermore z_1 is $(x_1 \oplus y_1) \oplus 1$ if $x_0 = y_0 = 1$ and z_1 is $(x_1 \oplus y_1) \oplus 0$ otherwise. In other words $z_1 = (x_1 \oplus y_1) \oplus (x_0 \wedge y_0)$. Bit z_2 is 1 if there is a carry from the second from the right addition or a propagation of a carry from the first (rightmost) addition with x_1 or $y_1 = 1$ to allow the propagation. Thus $z_2 = (x_1 \wedge y_1) \vee ((x_0 \wedge y_0) \wedge (x_1 \vee y_1))$. ■

Exercise 51.
Show that $P \vee \neg(P \wedge Q)$ is a tautology.

Solution.

$$\begin{aligned} P \vee \neg(P \wedge Q) &= P \vee (\neg P \vee \neg Q) \\ &= (P \vee \neg P) \vee \neg Q) \\ &= T \vee \neg Q \\ &= T \end{aligned}$$

■

Exercise 52.
Evaluate $\neg(P \vee Q) \vee (Q \wedge \neg P)$.

Solution.
One way to evaluate it is as follows

$$\begin{aligned}
\neg(P \vee Q) \vee (Q \wedge \neg P) &= (\neg P \wedge \neg Q) \vee (\neg P \wedge Q) \\
&= ((\neg P \wedge \neg Q) \vee \neg P) \wedge ((\neg P \wedge \neg Q) \vee Q) \\
&= ((\neg P \vee \neg P) \wedge (\neg Q \vee \neg P)) \wedge ((\neg P \vee Q) \wedge (\neg Q \vee Q)) \\
&= (T \wedge (\neg Q \vee \neg P)) \wedge ((\neg P \vee Q) \wedge T) \\
&= (\neg Q \vee \neg P) \wedge (\neg P \vee Q) \\
&= (\neg Q \vee \neg Q) \wedge \neg P \\
&= T \wedge \neg P \\
&= \neg P
\end{aligned}$$

There is a shorter way as follows

$$\begin{aligned}
\neg(P \vee Q) \vee (Q \wedge \neg P) &= (\neg P \wedge \neg Q) \vee (\neg P \wedge Q) \\
&= (\neg P \wedge (\neg Q \vee Q) \\
&= (\neg P \wedge T) \\
&= \neg P
\end{aligned}$$

■

2.3 Predicate calculus

Exercise 53.

(a) Let A be the population of all US citizens. Proposition $Q(x)$ is defined as follows.
$(\forall x \in A)(x$ **is named Alex**$)$.
We do not discuss the absurdity of this proposition. Find its negation $\neg Q(x)$.
(b) Let A be the population of all US citizens. Proposition $T(x)$, is defined as follows.
$(\exists x \in A)(x$ **is named Alex**$)$.
Quite clearly this proposition is true. Find its negation $\neg T(x)$.

Solution.
(a) Negation $\neg Q(x)$ is defined as follows: $(\exists x \in A)(x$ **is NOT named Alex**$)$. Quite clearly this proposition $\neg Q(x)$ is true, implying $Q(x)$ is false!
(b) Negation $\neq T(x)$ is defined as follows: $(\forall x \in A)(x$ **is NOT named Alex**$)$. The proposition $\neg T(x)$ might sound absurd and it is false, implying $T(x)$ is true.

■

Exercise 54.
Let P be "All students are in the classroom" State $\neg P$.

Solution.
Let S be the set of students referred to in P. Let C be the set of students in the classroom. We can rephrase P as follows

$$P : \forall s \in S : s \in C$$

The negation of P is $\neg P$ as follows

$$\neg P : \exists s \in S : s \notin C$$

Thus $\neg P$ is "there exists a student not in the classroom". ■

Exercise 55.
Let $A = \{1,3,5,7,9\}$. Determine the truth value of each of the following propositions.

- (a) P(n): $\exists n \in A : n+3 = 10$.
- (b) Q(n): $\exists n \in A : 2n+3 = 10$.
- (c) S(n): $\forall n \in A : 3n+7 \geq 10$.

Solution.
(a) T since for $n = 7$ we have $7+3 = 10$. Thus $T(P) = \{7\} \neq \emptyset$.
(b) F since $T(Q) = \{\}$.
(c) T since $T(S) = \{1,3,5,7,9\} = A$. ■

Exercise 56.
To disprove a $(\forall x)P(x)$ it suffices to find a so that $P(a)$ is F.
Is $(\forall x \in \mathbb{R})(\exists y \in \mathbb{Z})P(x,y)$ equivalent to $(\exists y \in \mathbb{Z})(\forall x \in \mathbb{R})P(x,y)$? Explain.

Solution.
Consider $P(x,y)$: y **is** $\lceil x \rceil$. $(\forall x \in \mathbb{R})(\exists y \in \mathbb{Z})P(x,y)$ reads every real number has an (integer) ceiling. However $(\exists y \in \mathbb{Z})(\forall x \in \mathbb{R})P(x,y)$ reads that there is an integer y that is the ceiling of all real numbers. This can't be true. ■

2.4 Elementary proof techniques

Exercise 57.

(Proof) by existence. For a proposition $P(x)$ show that $\exists x$ such that $P(x)$ is T.
For two real numbers a, b such that $a < b$, there exists a real number c such that $a < c < b$.

Solution.
We find an x that satisfies $a < x < b$. Let $x = (a+b)/2$. $x - a = (a+b)/2 - a = (b-a)/2 > 0$ and thus $x > a$.
Likewise $x - b = (a+b)/2 - b = (a-b)/2 < 0$ and thus $x < b$.
We have found c. It is $c = x = (a+b)/2$. ■

Exercise 58.

(Disproof) by counterexample. A theorem has a hypothesis and a conclusion. Find an instance x for which the hypothesis holds, and then show that for that instance x the conclusion DOES not hold. Or to prove that a proposition is false it suffices to provide a counterexample that show the proposition to be false.
Proposition Funny. Show that n^2+n+5 is a prime number for every integer n such that $n >= 0$.
A natural number (positive integer) p is prime if its only divisors are 1 and itself. A divisor a of p is an integer a that leaves a remainder of 0 when p is divided by a. Thus 1, and 5 are the only divisors of 5 since $5 = 5 \cdot 1$ and $5 = 1 \cdot 5$, and neither 2,3,4 divides 5. This is because $5 = 2 \cdot 2 + 1$, $5 = 3 \cdot 1 + 2$, $5 = 4 \cdot 1 + 1$, that is the division of 5 by 2,3,4 leaves a remainder of 1, 2, and 1 respectively: the corresponding quotients are 2, 1, and 1. Note that in $5 = 3 \cdot 1 + 2$, 5 is the divident, 3 is the divisor, 1 is the quotient, and 2 is the remainder of the integer division of 5 by 3.

Solution.
For $n = 4$, we have that $n^2+n+5 = 4^2+4+5 = 25$ and one divisor of 25 other than 1 and 25, is 5. Therefore 25 is not a prime number, it is in fact a composite number and therefore this simple counterexample shows Proposition Funny is false because it is not true for $n = 4$, where 4 is an integer. ■

Exercise 59.

Proposition Funny. Q(n): $\exists n \in \mathbb{N}$ such that n^2+7 is prime.
Show that Proposition Funny is T (true). Note: $\mathbb{N}$ is the set of natural numbers, $1,2,3,\ldots$ that is of positive integers. Sometimes it includes 0 and then we talk about non-negative integers.

Solution.
The existential quantifier makes things easy for us. In order to prove that Proposition Funny is true, we only need prove it for a single value of n. The existential quantifier reads 'there exist at least one natural number n'. Thus one natural number n is $n=3$ that make predicate Q(n) true. For $n=2$, we can easily establish that $2^2+7=11$ is a prime number. Proposition Funny is not, however, very interesting. Let us modify it a bit in the form of the problem to follow next. ■

Exercise 60.

Proposition NotFunny. $\forall n \in \mathbb{Z},\ n > 0\ n^2 + 7$ is prime.

Solution.
We can disprove this theorem for $n = 3$. $n^2 + 7 = 3^2 + 7 = 16$, and 16 is not prime. Its divisors other than 1, 16 include 2, 4, 8. ■

Exercise 61.

Proof by Direct Deduction. In order to prove $P \Rightarrow Q$, we 'Assume P (is true)', and using logical arguments 'derive Q (is true)'.
If integer n is such that $n > 0$ and n is even, then n^2 is even. Show it by direct deduction.

Solution.
We identify P as 'integer n, $n > 0$, and n is even'.

We identity Q as 'n^2 is even'.
STEP 1. Asume P: We assume n is integer, $n > 0$, and n is even.
STEP 2. For an even inger n, if we divide n by 2 the remainder of the division is a zero. Thus $n = 2k$ where k is the integer quotient of this division that leaves 0 as a remainder (i.e. $n = 2 * k + 0$).
Important conclusion-1: $n = 2k$.
Important conclusion-2: k is integer.
STEP 3. From step (2) above $n = 2k$. Subsequently, n^2 is equal to $n^2 = (2k)^2 = 4k^2$. This implies $n^2 = 2(2k^2) = 2m$, where $m = 2k^2$. In other words n^2 is a multiple of 2 and the other integer m. (m is integer since from (3) k is an integer and m which is $m = 2 * k^2 = 2 * k * k$ as the product of three integers is also an integer.
Important conclusion-3 $n^2 = 2m$.
Important conclusion-4 m is integer.
STEP 4. Conclusion $n^2 = 2m$, with m integer implies n^2 is even. The latter part is Q. ■

Exercise 62.
Show that for $n \in \mathbb{Z}$, $n \geq 1$ we have $n^2 + 5n + 12 \leq 18n^2$.

Solution.
We can prove this result directly. We use throughout this proof $n \in \mathbb{Z}$.
We have for all $n \geq 1$

$$12 \leq 12n^2$$

and

$$5n \leq 5n^2$$

Obviously

$$n^2 \leq n^2$$

If we add up the inequalities we have

$$n^2 + 5n + 12 \leq n^2 + 5n^2 + 12n^2 = 18n^2.$$

■

Exercise 63.

Proof by case analysis. Prove a theorem by case analysis. A theorem has a hypothesis and a conclusion.

- List all possible cases for which the hypothesis holds, and
- for each such possible case, do indeed prove that the conclusion holds.

Show that the product of two positive consecutive natural numbers $a, b \in \mathbb{N}$ is an even number.

Solution.
Let the first number be a. The next one (consecutive) would be $a+1$. Both are positive and thus $a > 0$.
Case analysis.
Case 1. a **is odd.** If a is odd then $a+1$ must be even. The product of an odd integer and an even integer is even.
Case 2. a **is even.** If a is even then $a+1$ must be odd. The product of an odd and even integer is even.
Conclusion: Both in case 1 and case 2 the product is an even number ■

Exercise 64.
Show that $\min(x,y)+\max(x,y)=x+y$.

Solution.
Proof by case analysis.
Case 1. Let $x<y$.
Then $\min(x,y)=x$. Also $\max(x,y)=y$. Therefore $\min(x,y)+\max(x,y)=x+y$.
Case 2. Let $x>y$.
Then $\min(x,y)=y$. Also $\max(x,y)=x$. Therefore $\min(x,y)+\max(x,y)=x+y$.
Case 3. Let $x=y$.
Then $\min(x,y)=y=x$. Also $\max(x,y)=x=y$. Therefore $\min(x,y)+\max(x,y)=x+y$. ■

Exercise 65.

Show that the square of a (positive) integer number retains the polarity of the integer number. The polarity of (positive) integer number x is 0 if it is even and 1 if it is odd.

Solution.
The square of a (positive) integer number retains the polarity of the integer number. Let x be a positive integer number. We are going to show that x and x^2 have the same polarity.
Case analysis.
Case 1. x **is odd.** If x is odd, the remainder of the division of x by two is 1. Thus $x = 2k+1$ where k is the quotient of the division. Then $x^2 = (2k+1)^2 = 4k^2+4k+1 = 4k(k+1)+1 = 2\cdot(2k(k+1))+1 = 2m+1$, where $m = 2k(k+1)$. $2m+1$ is an odd number. Thus if x is odd, then x^2 is also odd.
Case 2. x **is even.** If x is even, the remainder of the division of x by two is 0. Thus $x = 2k$ for some $k > 0$. Then $x^2 = 4k^2 = 2(2k^2) = 2m$ for some $m > 1$. Thus if x is even, then x^2 is also even.
Conclusion: Both in case 1 and case 2 the polarity of x is that of x^2. ■

Exercise 66.
For all A, B show that $A = (A \cap B) \cup (A - B)$.

Solution.
Direction 1: Show $A \subseteq (A \cap B) \cup (A - B)$.
Let $a \in A$. Then either $a \in B$ or $a \notin B$. If the former holds, we are done, since then $a \in A \cap B$ and then $a \in (A \cap B) \cup (A - B)$, since it belongs to the first component of the union. Otherwise consider the latter i.e. $a \notin B$. Then $a \notin A \cap B$. But then $a \in A - B$ since $a \notin B$. This implies $a \in (A \cap B) \cup (A - B)$ as well. Thus we have proved that an arbitraray $a \in A$ is also $a \in (A \cap B) \cup (A - B)$ and thus shown $A \subseteq (A \cap B) \cup (A - B)$.
Direction 2: Show $(A \cap B) \cup (A - B) \subseteq A$.
Let $x \in (A \cap B) \cup (A - B)$. Then either $x \in A \cap B$ or $x \in A - B$. This is because the intersection of the two sets $A \cap B$ and $A - B = A \cap B'$ is the $\emptyset$ since $A \cap B \cap A \cap B' = A \cap (B \cap B') = A \cap \emptyset = \emptyset$.
Case 1: $x \in A \cap B$. This implies that $x \in A$ (and also $x \in B$). However $x \in A$ leads to $x(A \cap B) \cup (A - B)$ leads to $x \in A$ and thus $(A \cap B) \cup (A - B) \subseteq A$ and we are done.
Case 2: $x \in (A - B)$. This implies again that $x \in A$, and we reach the same conclusion as in Case 1. Thus in both cases we have $(A \cap B) \cup (A - B) \subseteq A$.
Combining Direction 1 and Direction 2 we obtain the result. ■

Exercise 67.

Proof by contrapositive. The contrapositive of implication $P \Rightarrow Q$ is (implication) $\neg Q \Rightarrow \neg P$. In order to prove $P \Leftrightarrow Q$, we prove instead the contrapositive $\neg Q \Leftrightarrow \neg P$, through the direct method.
If positive integer $n > 0$ is odd, then show that n^2 is odd.

Solution.
We identify P as 'positive integer $n > 0$ is odd'.
We identity Q as 'n^2 is odd'.

Then $\neg P$ becomes 'positive integer $n > 0$ is even'.
Then $\neg Q$ becomes 'n^2 is even'.

Thus $P \Rightarrow Q$ becomes $\neg Q \Rightarrow \neg P$. The latter is proven by the direct method. Let n^2 be even. If n^2 is even then n must be even. (The latter is true since if n was odd $n = 2k+1$ and then $n^2 = (2k+1)^2 = 4k^2+4k+1 = 2(2k^2+2k)+1 = 2m+1$, would be odd too, that is impossible since n^2 is even.)
If $n > 0$ is odd, then n^2 is odd.
It suffices to prove that n^2 is NOT odd implies n is NOT odd.
Let n^2 be NOT odd. Then it must be even. If n^2 is even then n must be even i.e. n is NOT odd. Result proven. ■

Exercise 68.

Proof by contradiction. We show $P \Rightarrow Q$ by showing $P \wedge (\neg Q) \Rightarrow F$.
Show that the number of prime number is infinite, i.e. there is no largest prime number in other words there infinitely many.

Solution.
Let P : be let $p_1, \ldots, p_j$ be the prime numbers less than or equal to p_j. Let Q : be $\exists q$, such that $q > p_j$ and q prime. We are going to prove the result by contradiction i.e. $P \wedge (\neq Q) \Rightarrow F$.
$\neg Q$ is equivalent to the fact that $p_1, \ldots, p_j$ is the only set of prime numbers i.e. they are finite. Form $n = p_1 \cdot p_2 \cdot \ldots \cdot p_j + 1$. The remainder of the division of n by p_i is 1 thus no p_i divides n. There are two possibilities n is a prime number contradicting $\neg Q$ or n is a composite divisible by a prime number q other than p_i. The latter also contradicts $\neg Q$. ■

Exercise 69.

Proof by contradiction. We show $P \Rightarrow Q$ by showing $P \wedge (\neg Q) \Rightarrow F$. F can be $(\neq Q) \wedge Q$ or not!
Show that $\sqrt{2}$ is (an) irrational number. (Be reminded that x is rational if and only if there are integer $m, n \neq 0$ such that $x = m/n$.)

Solution.
Let us assume that $\sqrt{2}$ is rational that is $\sqrt{2} \in \mathbb{Q}$. We are going to show that this leads to a contradiction.
If $\sqrt{2}$ is rational then there exist $m, n \in \mathbb{Z}$, $m, n \neq 0$ such that $\sqrt{2} = m/n$. **Integer m and n have no common divisor**; otherwise we can divide it out from m, n, and reach m' and n' such that $\sqrt{2} = m'/n'$ and m' and n' share no common divisor.
Then squaring both sides we have

$$(\sqrt{2})^2 = (m/n)^2 \iff 2 = m^2/n^2 \iff m^2 = 2n^2.$$

We thus have that m^2 is an even number since it is of the form $2k$, with $k = n^2$. Then, by the previous theorem, m **must also be even**.
If m is even then $m = 2p$ for some integer p.

$$m = 2p \iff m^2 = (2p)^2 \iff m^2 = 4p^2$$

But $m^2 = 2n^2$. Therefore

$$m = 2p \iff m^2 = 4p^2 \iff 2n^2 = 4p^2 \iff n^2 = 2p^2.$$

Thus n^2 is an even number and by a previous theorem, n **must also be even**.
Since n and m are even, then they have a common divisor which is two. This contradicts the choice of m, n that were picked so that they share NO COMMON DIVISOR. ■

Exercise 70.
What is the negation of the implication: if A then B ($A \Rightarrow B$).

Solution.
$A \Rightarrow B$ is equivalent to $(\neg A) \vee B$. Then $(\neg((\neg A) \vee B))$ is $A \wedge (\neg B)$, i.e. A and not B. Note that the contrapositive of $A \Rightarrow B$ is $\neq B \Rightarrow \neg A$.
Thus if $A \Rightarrow B$ is false, its negation is true and thus $A \wedge (\neg B)$ is true. The contrapositive $\neq B \Rightarrow \neg A$ is equivalent to $B \vee \neg A$. But the latter$B \vee \neg A$ and $A \wedge (\neg B)$ are of opposite polarity, a contradiction. ■

Exercise 71.

Proof by converse. Let P and Q be two propositions. The converse of implication $P \Rightarrow Q$ is (implication) $Q \Rightarrow P$.
(a) Let I be the implication: If $x \leq y$ and $x \neq y$ then $x < y$.
The converse of I is the following implication: If $x < y$ then $x \leq y$ and $x \neq y$.
(b) If triangle ABC is isosceles then AB=AC. Conversely, if AB=AC then triangle ABC is isosceles.

Solution.
Obvious to prove (both). ■

Exercise 72.
An integer $n > 1$ is prime if its only positive divisor other than 1 s the integer n itself. Otherwise integer n is composite. Examine the truth value of the proposition P defined as follows:

$$P : 15 \textbf{ is prime}$$

Solution.
Proposition P is F. We can prove $\neg P$ is T and thus P is F. $\neg P$ is *15 is NOT prime* or equivalently *15 is composite*. We know that $15 = 3 \cdot 5$ thus 3 divides 15 and 5 divides 15. 15 is indeed composite. Neither 3 nor 5 is 1 or 15 itself.

■

Exercise 73.
Show that $7^n - 1$ is a multiple of 6.

Solution.
Use the identity $a^n - b^n = (a-1)(a^{n-1} + a^{n-2} + \ldots + a + 1)$ Then $7^n - 1^n = 6(7^{n-1} + \ldots + 7 + 1) = 6k$. Thus $7^n - 1$ is a multiple of 6. ■

Exercise 74.
If n,m are odd integers, then $n+m$ is an even integer.

Solution.
Direct proof. Let n be odd. This means there exists integer k such that $n = 2k+1$, i.e. the remainder of the division of n by 2 is 1. Likewise m is odd and thus there exists integer l such tham $m = 2l+1$ for the same reason. Then $n+m = 2k+1+2l+1 = 2(k+l+1)$. Thus $n+m$ is a multiple of 2 i.e. an even number. ■

Exercise 75.
If n is an odd integer and m is an even integer, then $n+m$ is an odd integer.

Solution.
See previous problem. ■

Exercise 76.
If n is odd, so is n^2. Use a direct proof to prove it.

Solution.
Direct proof. Let n be odd. This means there exists integer k such that $n = 2k+1$, i.e. the remainder of the division of n by 2 is 1. Then $n^2 = (2k+1)^2 = 4k^2+4k+1 = 4k(k+1)+1 = 2(2k(k+1))+1 = 2l+1$, where $l = 2k(k+1)$ is an integer as the product of three integers: 2, k and $k+1$. Because $n^2 = 2l+1$, we conclude n^2 is an odd number. ■

Exercise 77.
If n is odd, so is n^2. Use a proof by contradiction to prove it.

Solution.
Proof by contradiction.
Let n be odd. This means there exists integer k such that $n = 2k+1$, i.e. the remainder of the division of n by 2 is 1. Then $n^2 = (2k+1)^2 = 4k^2+4k+1 = 4k(k+1)+1 = 2(2k(k+1))+1 = 2l+1$, where $l = 2k(k+1)$ is an integer as the product of three integers: 2, k and $k+1$.
We would like to prove that n^2 is also odd. Let, for the sake of a contradiction, that n^2 is even.
Based on the prior analysis, we concluded that n^2 is of the form $n^2 = 2l+1$, i.e. n^2 is an odd number that contradicts the assumption that n^2 is even. Thus n^2 cannot be even. Conclusion: n^2 is odd. ■

Exercise 78.

Show that the product of two consecutive integers is even.

Solution.
Let n and $n+1$ be two consecutive integers. We do a case analysis.
Case 1: n is odd. Then, as previously noted, there exists integer k such that $n = 2k+1$. Then $n+1 = 2k+2$ and thus $n(n+1) = (2k+1)(2k+2) = 2(k+1)(2k+1) = 2l$ with l an integer such that $l = (k+1)(2k+1)$. Thus $n(n+1) = 2l$ is a multiple of two i.e. even.

Case 2: n is even. Then, as previously noted, there exists integer k such that $n = 2k$. Then $n+1 = 2k+1$. Similarly as before $n(n+1) = 2(k(2k+1))$ we conclude $n(n+1)$ is even as well.
Whethere n is odd or even, $n(n+1)$ is even. ■

Exercise 79.

Show that the square of an integer n is either a multiple of 4 or leaves a remainder of 1 if divided by 8. For example $2^2, 6^2, 8^2, 10^2$ are all multiples of 4. $3^2, 5^2, 7^2, 9^2$ are $8+1$ or $3\cdot 8+1$, or $6\cdot 8+1$, or $10\cdot 8+1$ respectively.

Solution.
We do a case analysis.
Case 1: n is even. Then (skipping some details obtainable from previous exercises), $n = 2k$ and thus $n^2 = 2k \cdot 2k = 4k^2$ i.e. n^2 is a multiple of 4

Case 2: n is odd. Then $n = 2k+1$ and thus $n^2 = (2k+1)^2 = 4k(k+1)+1$. But we know the product of two consecutive integers is even thus $k(k+1) = 2l$. Then $n^2 = 4k(k+1)+1 = 4(2l)+1 = 8l+1$. Thus the remainder of the division of n^2 by 8 is 1 as shown (and the uniqueness of integer division).
With Case 1 (square multiple of 4) and Case 2 (square leaves remainder 1 after division with 8) we complete the proof. ■

Exercise 80.
A three digit denary (base-10) integer is divisible by (multiple of) 3 if the sum of its digits is divisible by 3. Thus 123 is a multiple of 3 since $1+2+3=6=3\cdot 2$. And 972 is a multiple of 3 since $9+7+2=3\cdot 6$. Thus 123 is a multiple of 3 since $1+2+3=6=3\cdot 2$. And 972 is a multiple of 3 since $9+7+2=3\cdot 6$. (By the way $123=3\cdot 41$ and $972=3\cdot 324$.)

Solution.
Let $n=abc$ be a three digit denary integer. Then a is the number of hunders, b the number of tens and c the units of n. Thus $n=a\times 100+b\times 10+c$. We know $100=3\cdot 33+1=3k+1$ with $k=33$. We know $10=3\cdot 3+1=3l+1$ with $l=3$. Tehn $n=a(3k+1)+b(3l+1)+c=3(ak+bl)+(a+b+c)$. If $a+b+c$ is divisible by 3, then $a+b+c=3q$ for some integer q. Then $n=3(ak+bl)+(a+b+c)=3(ak+bl+q)$ and thus n is a multiple of 3. ■

Exercise 81.
Any three digit integer that ends with 5 has a square that is a multiple of 25. Thus $ab5$ is such that $(ab5)^2$ is a multiple of 25.

Solution.
Let $ab5 = a \times 100 + b \times 10 + 5 = 10(10a+b)+5$. Then, $(ab5)^2 = 100(10a+b)^2 + 25 + 100(10a+b) = 25k$ where $k = 4(10a+b)^2 + 1 + 4(10a+b)$. This concludes the proof. ■

Exercise 82.

Prove by contradiction that for all $x, y \in \mathbb{R}$ with $x + y \geq 100$, then $x \geq 50$ or $y \geq 50$.

Solution.
We want to prove the implication.

$$\forall x \forall y : x + y \geq 100 \Rightarrow x \geq 50 \vee y \geq 50$$

Let P be the proposition of the antecedent.

$$P : \forall x \forall y : x + y \geq 100$$

Let Q be the proposition of the consequent

$$Q : x \geq 50 \vee y \geq 50$$

Thus the implication states

$$P \Rightarrow Q$$

The negation $\neg Q$ of Q is

$$\neg Q : x < 50 \wedge y < 50$$

Contradiction: Assume P is true but Q is F.
Thus for the sake of contradiction we are going to assume $\neg Q$ is T or equivalently Q is F. This means that both x, y are less than 50. That is $x < 50$ and $y < 50$. Then $x + y < 50 + 50 = 100$ i.e. $x + y < 100$.
This however contradicts the antecedent of the implication which is proposition P that states that for all x, y that $x + y \leq 100$.
We have thus proven that P is false. This contradicts the FACT (assumption) that P is true.
In other words if Q is F we proved because of $x + y \leq 100$ a contradiction to the antecedent's $x + y \geq 100$.
Thus Q cannot be false. Thus Q must be true. ■

Exercise 83.

Show by contrapositive that for all $x \in \mathbb{R}$ if x^2 is irrational then x is irrational.

Solution.
We want to prove the implication.

$$\forall x : x^2 \textbf{ is irrational } \Rightarrow x \textbf{ is irrational}$$

Let P be the proposition of the antecedent.

$$P : \forall x : x^2 \textbf{ is irrational}$$

Let Q be the proposition of the consequent

$$Q : x \textbf{ is irrational}$$

Thus the implication states

$$P \Rightarrow Q$$

Proof by contrapositive means

$$\neg Q \Rightarrow \neg P$$

That is we show that if x is rational then x^2 is also rational. If x is rational there exist integer n, m such that $x = n/m$. Then $x^2 = n^2/m^2$. If n, m are integer so are n^2, m^2 as the products of two integers. Then x^2 is rational as the fraction of two integers.
Proof is complete because we proved $\neg Q \Rightarrow \neg P$. By contrapositive, $P \Rightarrow Q$ has been proven. ■

Exercise 84.
Show the following: there are no positive integers (strictly) between 0 and 1.

Solution.
The proposition is true. Say there is at least one integer a strictly between 0 and 1 i.e. $0 < a < 1$. Collect a and all those like a (integers strictly beetween 0 and 1) into set A. We have $|A| \geq 1$ since $a \in A$. By the Well Ordered Set Principle A has a minimal element. Call it m. Square m to obtain m^2. Since $0 < m < 1$ we have that $0 < m^2 < 1$. Thus m^2 is also a member of A. Moreover, $0 < m^2 < m < 1$. That is m^2 is an element of A smaller than its smallest element A which is m! This contradicts the minimality of m. Thus a cannot exist which means A is empty! ■

Exercise 85.
Show that there is no natural number between natural number k and natural number $k+1$, the successor of k, where $k \in \mathbb{N}$.

Solution.
If $k = 0$ the result was proven in the prior problem. Let t be an integer between k and $k+1$ for the case $k \neq 0$. Then

$$k < t < k+1$$

Consider $m = t-k$. It is clear that

$$0 < t-k < 1$$

Moreover $t-k$ is the difference between two natural numbers. Therefore it is a natural number. By the previous problem such $m = t-k$ does not exist. ■

2.5 Peano-based proof methods: induction

Exercise 86.
Show that $\sqrt{2}$ is irrational.

Solution.
Say that $\sqrt{2}$ were rational. Then $\sqrt{2} = p/q$ for some integers p,q. Form then set A whose elements are $r\sqrt{2}$ where both r and $r\sqrt{2}$ are integer. Set A contains p since $p = q\sqrt{2}$, p,q are integer and thus $p \in A$ $(r = q)$. By the Well Ordered Set Principle, A has a minimum and let it be $M = m\sqrt{2}$ (with M,m both integer). Since $\sqrt{2} > 1$ we have that

$$M(\sqrt{2}-1) = M\sqrt{2} - M = M\sqrt{2} - m\sqrt{2} = (M-m)\sqrt{2}$$

is an integer: $M = m\sqrt{2}$ is an integer by assumption, and $M\sqrt{2} = 2m$ also an integer. We the obtain

$$(M-m)\sqrt{2} = (m\sqrt{2}-m)\sqrt{2} = m(2-\sqrt{2}) < m\sqrt{2} = M$$

We use $2-\sqrt{2} < \sqrt{2}$ since $\sqrt{2} > 1$. The $(M-m)\sqrt{2}$ is (a) integer, and (b) $M-m$ is also an integer i.e. a member of A. This contradicts the mimimality of M since we found $(M-m)\sqrt{2}$ smaller than M. ■

Exercise 87.
Show that Theorem 2.2 (weak form of induction) is TRUE.

Solution.
Let $A^{\complement}$ be the complement of A over $\mathbb{N}$.

$$A^{\complement} = N - A = \{k \in N | k \notin A\}.$$

Suppose $A \neq N$. Then $A^{\complement}$ is a non-empty set and by the Well ordered set principle it has a minimal element and call that element a. Since $0 \in A$ it is obviously $a \neq 0$. This mean $a > 0$ and thus $a-1$ is non-negative (positive or zero). Because a is the mimimal element of $A^{\complement}$, $a-1$ belongs to A. Thus $a-1 \in A$. From the statement of the mathematical induction principle set $k = a-1$. $k \in A$ implies $k+1 = a \in A$. This contradicts the membership of a in the complement of A! Thus $A^{\complement}$ cannot have a minimal element i.e. it must be empty. This implies that $A = N$.
The argument works also if the definition of $\mathbb{N}$ is the set of positive integers thus exluding zero.
Reiterating: If a set A contains k, and also contain $n+1$ whenever it contains n, where $n > k$, then A contains all the integers greater than or equal to k. ■

Exercise 88.

Proof by (ordinary, or weak) induction. The Well Ordered Set Principle can be used to prove that

'P(n) is true $\forall n \in \mathbb{N}$' .

This can be done as follows.

1. Formulate a set C of counterexamples to $P(n)$ being true, thus defining

$$C = \{n \in N : \neg P(n)\}$$

(This reads 'C contains all integers n such that $\neg P(n)$ is true'. This latter '$\neg P(n)$ is true' implies '$P(n)$ is false'.)

2. Assume (for the sake of contradiction) that C is non-empty.
3. By the Well Ordered Set Principle C has a minimal element m.
4. Obtain a contradiction by finding a $c \in C$ smaller than m.
5. Then conclude that C must be empty i.e. $P(n)$ is true for all n.

Use the Well Ordered Set Principle to prove the following result: Prove by induction that $1+2+\ldots+n = n(n+1)/2$ for all $n \in \mathbb{N}$.

Solution.
(a) **Formal way of showing a Proposition by induction.** By way of contradiction, assume that the theorem is false. Then there exist some integers that serve as counterexamples and thus populate C.

1. Form C
$$C = \{k \in N : 1+2+\ldots+k \neq k(k+1)/2\}$$
2. If C is empty we are done. If C is non-empty then it has a minimal element by the W.O.S.P. and let that be m.
3. Thus for m we have $P(m)$ is not true i.e. $P(m)$ is false and thus
$$1+2+\ldots+m \neq m(m+1)/2$$
Consider $k = 1$. And also $k = 0$.
$$1 = 1(1+1)/2$$
$$0 = 0(0+1)/2$$
Thus $0 \notin C$. Also $1 \notin C$. Therefore $m > 1$. Consider $m-1 < m$. Then $P(m-1)$ is true since m is smallest value for which $P(m)$ is false. Then $P(m-1)$ true implies.
$$1+2+\ldots+(m-1) = (m-1)m/2.$$
If we add m to both sides, we obtain.
$$1+2+\ldots+(m-1)+m = (m-1)m/2+m = m(m+1)/2.$$
This contradicts
$$1+2+\ldots+m \neq m(m+1)/2$$
4. Consequently C cannot be non-empty.
5. Therefore, C must be empty, that is $P(n)$ is true for all $n \in \mathbb{N}$.

We can prove also that Every non-empty finite subset of real numbers has a minimal element. ■

Exercise 89.

We prove the result of the previous problem less formally now.

Induction as a proof method. In inductive proofs we try to prove that the set of integers A that satisfy a given property or proposition is all of $\mathbb{N}$ (or $\mathbb{N}^*$) i.e. $A = \mathbb{N}$. (We drop the alternative form from now on.) In the remainder we ignore properties and focus on propositions. Of particular interest are propositions that depend on an integer variable n and thus we would establish that the range A of n is indeed $\mathbb{N}$ i.e. $A = \mathbb{N}$.

Let for any natural number $k > 0$, $S_k(b)$ be defined as follows.

$$S_k(n) = 1^k + 2^k + \ldots + n^k$$

We also write $S_k(n) = \sum_{i=1}^{n} i^k$. Prove by induction that $A(n) = S_1(n) = 1 + 2 + \ldots + n$ is equal to $n(n+1)/2$ for all $n \in \mathbb{N}$ using induction.

Solution.
STEP 0: Identify the predicate $P(n)$ that depends on an integer/natural variable, and also that variable.
The first step in induction is to identify in a proposition a predicate that depends on a natural-valued variable and also that same natural-valued variable. Obviously the predicate $P(n)$ is $A(n) = n(n+1)/2$, where $A(n)$ is another name for the generic $S_1(n)$. We are going to show that $P(n)$ is true for all (integer $n \geq 1$). That is we shall show that $A(n) = n(n+1)/2$. The proof is by induction.
STEP 1: Base case: (Show that) $P(1)$ is true.

The left hand-side of the equality in $P(n)$ is $A(n)$ i.e. $S_1(n)$ and the right-hand side is $n(n+1)/2$. We shall show that the two sides are equal for $n = 1$ i.e. we shall show $P(1)$.
Left hand side of $P(n)$ first for $n = 1$: $A(n)|_{n=1} = 1$. The left hand side sum of $P(1)$ is $A(1)$ i.e. the sum of one term (and that term is 1), which is 1.
Right hand side of $P(n)$ for $n = 1$: $n(n+1)/2_{n=1} = 1(1+1)/2 = 1$. Obviously this is $1(1+1)/2 = 1$.
Therefore $P(1)$ is true since the left hand side $A(1)$ is equal to the right hand side $1(1+1)/2$.
STEP 2: Inductive Step. $P(n) \Rightarrow P(n+1)$.

2.a Induction hypothesis: $P(n)$. A $P(n)$ true is equivalent to

$$A(n) = 1 + 2 + \ldots + n = \sum_{i=1}^{n} i = n(n+1)/2.$$

2.b How do we establish $P(n+1)$ then? A $P(n+1)$ (to be proven true) is equivalent to

$$A(n+1) = 1 + 2 + \ldots + (n+1) = \sum_{i=1}^{n+1} i = (n+1)(n+2)/2.$$

2.c Use 2.a to get to 2.b.

Starting with 2.b we write down the left hand side of $A(n+1)$, and then we separate the last term from the

previous n terms of the sum

$$\begin{aligned}
A(n+1) &= \sum_{i=1}^{n+1} i \\
&= 1+2+\ldots+(n+1) = (1+2+\ldots+n)+(n+1) \\
&= \underbrace{\left(\sum_{i=1}^{n} i\right)}_{A(n)} +(n+1) \\
&= \underbrace{\left(\sum_{i=1}^{n} i\right)}_{Use\ P(n)} +(n+1) \\
&= n(n+1)/2+(n+1) = n(n+1)/2+2(n+1)/2 = (n+1)(n+2)/2
\end{aligned}$$

This completes the inductive proof. We proved two things

- We first proved that $P(1)$ is true.
- and then showed $P(n) \Rightarrow P(n+1)$ for all $n \geq 1$.

■

Exercise 90.

In inductive proofs we try to prove that the set of integers A that satisfy a given property or proposition or predicate P is all of $\mathbb{N}$ i.e. $A = \mathbb{N}$. We relax this and use $\mathbb{N}^*$ instead, where $\mathbb{N}^* = \mathbb{N} - \{0\}$.
(i) Show the following using induction. If $a, b \in \mathbb{N}^*$ then $ab \geq a$. Equality is applicable if and only if $b = 1$.

Solution.
(i) Use (alternative ordinary) induction on b. Let $a \in \mathbb{N}^*$.
STEP1: Base case. It is $b = 1$. Clearly $a \cdot 1 \geq a$ since $a \cdot 1 = a$.

Auxiliary step. For the inductive step, since $a \in \mathbb{N}^*$, we have $a \geq 1$, which implies $a > 0$ and thus $2a = a + a > a + 0$, leads to $2a > a$.
STEP $1'$: Inductive hypothesis. Suppose that $ab > a$ for some $b \in \mathbb{N}$. We then will show that $a(b+1) > a$ in the inductive step. The inductive step is formulated below.
STEP 2: Inductive step.

$$ab > a, \text{ for some } b \in \mathbb{N} \implies a(b+1) > a.$$

By the inductive hypothesis $ab > a$. Therefore

$$a(b+1) = ab + a > a + a = 2a,$$

i.e. $a(b+1) > 2a$. Moreover by the auxiliary step $2a > a$. By way of trasitivity $a(b+1) > 2a > a$ and the inductive step has been shown.
STEP 3: Conclusion. It follows that $ab > a$ for all $b \geq 2$. For the base case $b = 1$ we have $ab \geq a$ in fact $ab = a$. ■

Exercise 91.

Is the inductive proof of the Proposition below correct? Explain.

> **Proposition 2.4**
>
> Proposition: All horses of the world are of the same color.

Solution.
The proof is (supposed to be) by induction on the number of horses n.

P(n): In any set of $n \geq 1$ horses, all the horses of the set are of the same color.

1. Base Case: Show $P(1)$ is true. $P(1)$ is always true as in a set consisting of a single horse, all the horses (there is only one) of the set have the same color.

2. Inductive step : $\forall n \in N$, i.e. $n \geq 1$, $P(n) \Rightarrow P(n+1)$. Let us assume (induction hypothesis) that for any $n \geq 1$, $P(n)$ is true. Since we assume $P(n)$ to be true, every set of n horses have the same color. Then we will prove that $P(n+1)$ is also true (inductive step), i.e. we will show that in every set of $n+1$ horses, all of them are of the same color. To show the inductive step, i.e. that $P(n+1)$ is true let us consider ANY set of $n+1$ horses $H_1, H_2, \ldots H_n, H_{n+1}$. The set of horses $H_1, H_2, \ldots, H_n$, consists of n horses, and by the induction hypothesis any set of n horses are of the same color. Therefore color(H_1) =color$(H_2) = \ldots =$color(H_n). The set of horses $H_2, H_3, \ldots, H_{n+1}$, consists of n horses, and by the induction hypothesis any set of n horses are of the same color. Therefore color(H_2) =color$(H_3) = \ldots =$color(H_{n+1}). Since from the first set of horses color(H_2) =color(H_n), and from the second set color(H_2) =color(H_{n+1}), we conclude that the color of horse H_{n+1} is that of horse H_2, and since all horses $H_1, H_2, \ldots, H_n$ are of the same color, then all horses $H_1, H_2, \ldots, H_n, H_{n+1}$ have the same color. This proves the inductive step. The induction is complete and we have thus proved that for any n, in any set of n horses all horses (in that set) are of the same color. ■

Exercise 92.

What is wrong with the proof of Proposition 2.4? Find it and reason about it.

Solution.
Horses revisited i.e. why the Theorem of Example 2.4 is false. On the previous page we proved that all horses have the same color, an otherwise nonsense statement (or false proposition). What's wrong with the inductive proof?
Many may argue that the error is in the logic of the inductive step.
The logic is fine, the quantification "for all n" is not, since the assumption of always having horse H_2 might not be true. The crux of the inductive step is the existence of three 'different' horses H_1, H_2, H_{n+1}. We first form a set of n horses $H_1, H_2, \ldots, H_n$ and apply the induction hypothesis and then form another set of n horses, $H_2, \ldots H_n, H_{n+1}$, and apply the induction hypothesis again. Crucial to the proof is that $c(H_1) = c(H_2)$ from the first application of the induction hypothesis, and $c(H_2) = c(H_{n+1})$ from the second thus concluding that $c(H_1) = c(H_2) = \ldots = c(H_n)$.
Let's see what happens for $n = 1$, i.e. let's try to show that $P(1) \Rightarrow P(2)$, i.e. show the inductive step for a certain value of n equal to 1. If we try to form $H_1, \ldots H_n$ this set contains only one 'horse' element for $n = 1$: H_1 ! If we try to form $H_2, \ldots, H_{n+1}$ this set contains only one 'horse' element H_2. There is no common third 'horse' H_k in the set containing H_1 nor in the set containing H_2. This is because the argument in the previous paragraph works only for $n \geq 2$. In that case one set is formed from H_1, H_2 and the other set from H_2, H_3. However even if we can prove the inductive step nicely in that case there is no way to prove the base case set for $n = 2$ i.e. $P(2)$!
Therefore the inductive step that "we proved" before $P(n) \Rightarrow P(n+1)$ is not true for all $n \geq 1$ but only for $n \geq 2$. This however can not establish the trueness of $P(n)$ for all $n \geq 1$ because $P(2)$ may or may not be true.
What is $P(2)$?

$P(2)$ **is "in any set of two horses, both horses are of the same color".**

In conclusion, the whole "horsy argument" breaks down because
A2. $P(n) \Rightarrow P(n+1)$ for all $n \geq 1$,

was not shown, for all $n \geq 1$; it was only proved for all $n \geq 2$, i.e.
A2. $P(n) \Rightarrow P(n+1)$ for all $n \geq 2$,

and thus the base case "$P(1)$ is true" can not be used with the latter version of the inductive step; for the latter we need $P(2)$ to be true WHICH IS NOT! ■

Exercise 93.

Let for any natural number $n \geq 0$ and $x \in R$ we define

$$G(n,x) = \sum_{i=0}^{n} x^i = x^0 + x^1 + \ldots + x^i + \ldots + x^n.$$

The sum is known as the geometric series or sum of the terms of the geometric sequence. The i-th term of the sum is x^i.

Let $P(n)$ be defined as follows. $P(n) : \forall n \geq 0, x \neq 1$,

$$G(n,x)(= \sum_{i=0}^{n} x^i) = \frac{x^{n+1}-1}{x-1}.$$

Show that $P(n)$ is true for all $n \geq 0$ (and $x \neq 1$).

Solution.
We prove the theorem by induction.
STEP 0. Identify predicate $P(n)$. The predicate $P(n)$ that depends on n is given to us

$$P(n) : G(n,x) = \frac{x^{n+1}-1}{x-1},$$

where $G(n,x) = \sum_{i=0}^{n} x^i$.
STEP 1. Base case: (Show that) $P(0)$ is true. We can show either $P(0)$ or $P(1)$: then base case would be $n = 0$ or $n = 1$. We show $P(0)$ is true i.e. we use $n = 0$.
The left-hand side of (the equality in) $P(n)$ $G(0,x) = \sum_{i=0}^{0} x^i = x^0 = 1$.
The right-hand side is $\frac{x^{n+1}-1}{x-1}|_{x=1} = \frac{x^1-1}{x-1} = 1$.
Right hand-side is equal to left-hand side. Therefore $P(1)$ is true.
STEP 2. Inductive Step: $P(n) \Rightarrow P(n+1)$.

2.a Induction hypothesis: $P(n)$. A $P(n)$ true leads to

$$G(n,x)(= 1 + x + \ldots + x^n) = \frac{x^{n+1}-1}{x-1}.$$

2.b How do we establishe $P(n+1)$ then? A $P(n+1)$ that is to be proven true is equivalent to

$$G(n+1,x)(= 1 + x + \ldots + x^{n+1}) = \frac{x^{n+2}-1}{x-1}.$$

We start from the latter's $G(n+1,x)$ left-hand side to conclude its right hand-side by using the former result for $P(n)$. Between the second and third inequality we use the induction hypothesis for $P(n)$ above.
2.c Use 2.a to get to 2.b.

$$\begin{aligned} G(n+1,x) &= 1 + x + \ldots + x^{n+1} = 1 + x + \ldots + x^n + x^{n+1} \\ &= (1 + x + \ldots + x^n) + x^{n+1} \\ &= \frac{x^{n+1}-1}{x-1} + x^{n+1} \\ &= \frac{(x^{n+1}-1) + x^{n+1}(x-1)}{x-1} \\ &= \frac{x^{n+2}-1}{x-1}, \end{aligned}$$

which establishes $P(n+1)$ from $P(n)$. Base case and inductive step combined conclude that $P(n)$ is true, as needed. ■

Exercise 94.

Show by induction that for all integer $n \geq 1$, we have that $1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1) \leq n^n$.

Solution.
Base case $n = 1$**.** It is clear that $1 \leq 1^1$.
Induction step. Let the inequality be true for $n = k$ that is

$$1 \cdot 3 \cdot \ldots \cdot (2k-1) \leq k^k.$$

We will show that it is true for $n = k+1$ that is

$$1 \cdot 3 \cdot \ldots \cdot (2k-1)(2k+1) \leq (k+1)^{k+1}.$$

$$\begin{aligned} 1 \cdot 3 \cdot \ldots \cdot (2k-1)(2k+1) &\leq (1 \cdot 3 \cdot \ldots \cdot (2k-1)) \cdot (2k+1) \\ &\leq k^k \cdot (2k+1) \\ &\leq (k+1)^{k+1} \end{aligned}$$

In order to show that $k^k \cdot (2k+2) \leq (k+1)^{k+1}$ we start with

$$\begin{aligned} \frac{(k+1)^{k+1}}{k^k} &\geq (k+1) \cdot (1+1/k)^k \\ &\geq (k+1) \cdot 2 \\ &\geq 2k+1. \end{aligned}$$

As far as the proof that $(1+1/k)^k \geq 2$ is concerned, by the binomial theorem

$$(1/k+1)^k = \sum_{l=0}^{k} \binom{k}{l} (1/k)^l 1^{k-l} = 1+1+\sum_{l=2}^{k} \binom{k}{l} (1/k)^l 1^{k-l} \geq 2.$$

If the binomial theorem is not available use a simplified form of the Bernoulli theorem that states that for all $x > -1$ and positive integer n the following is true.

$$(1+x)^n \geq 1+nx$$

A simple proof for $x > 0$ is by induction on n. For $n = 1$ it is obviously true. If it is true for n, the

$$(1+x)^{n+1} = (1+x)^n(1+x) \geq (1+nx)(1+x) = 1+(n+1)x+nx^2 \geq 1+(n+1)x.$$

■

Exercise 95.
Prove by induction that $2^n > (n+1)^2$ for all $n \geq 6$.
Prove also that for $n \geq 64$ we have that $\sqrt{n} > \lg n$.

Solution.
For $n = 6$ we need to show $2^6 > (6+1)^2$ which is true thus concluding with the base case.
For the inductive step $2^n > (n+1)^2$ implies to $2^{n+1} > (n+2)^2$.

$$\begin{aligned} 2^{n+1} &> 2 \cdot 2^n \\ &> 2 \cdot (n+1)^2 \\ &> 2 \cdot (n^2 + 2n + 1) \\ &= 2n^2 + 4n + 2 \\ &= n^2 + 4n + 4 + n^2 - 2 \\ &= (n+2)^2 n^2 - 2 \\ &> (n+2)^2 \end{aligned}$$

The latter is a result of $n^2 - 2 > 0$ for $n \geq 6$ and in fact for $n \geq 2$.
Omitted. ■

Exercise 96.
Any integer $n > 1$ has a unique factorization into prime numbers. Thus if $n = p_1 p_2 \dots p_k$ is one factorization with $p_i \leq p_j$ for $i < j$, and $n = q_1 q_2 \dots q_l$ is another factorization with $q_i \leq q_j$ for $i < j$, then $k = l$ and $p_i = q_i$ for all i.

Solution.
This material assume knowledge of divisibility.
For $n > 1$ the result is true for $n = 2$ since integer 2 is a factorization of itself and 2 is a prime number. Let us assume that every integer between 2 and n inclusive has a unique factorization. We are going to show the same for $n+1$. If $n+1$ is a prime number, we are done. Otherwise it is the product of two number smaller than $n+1$. Let $n+1 = ab$. Then $a, b < n+1$ that is $a, b \leq n$, and by the inductive hypothesis a, b have unique factorizations. Let them be $a = r_1 r_2 \dots r_k$ and $a = s_1 s_2 \dots s_l$. Then $n+1 = r_1 r_2 \dots r_k s_1 s_2 \dots s_l$ and therefore $n+1$ has a factorization. We need to show it is a unique on.
Say some integer has two or more factorizations that are different. If there are more than one such integers pick the smallest one and call it n. If there are more than two factorizations, pick any two of them and let them be as follows. $n = p_1 p_2 \dots p_k$ and $n = q_1 q_2 \dots q_l$. We have that p_1 divides n. Then p_1 divides $q_1 q_2 \dots q_l$. Therefore since p_1 is a prime number and so are all q_i we have that $p_1 = q_j$ for some j. Divide the two forms of n (as a product of p_is and a product of q_is) with $p_1 = q_j$. We get a smaller number than n that has two different factorizations. (A common term was cancelled out that was common in the two factorizations of n that in total were different.) The smaller number cannot have two different factorizations because the smalle number with two or more different factorizations was n. (If the smaller number has a unique factorization, adding to it p_1 or q_j generate another unique factorization.) ■

2.6 Miscellaneous problems

Exercise 97.
Let $A, B \neq 0$ be two integers. We define Q, R the quotient and the remainder of the integer division of A by B. This makes A the dividend and B the divisor.

$$A = B \cdot Q + R.$$

For the division to be uniquely definesd we impose a condition that $0 \leq R < |B|$.
(a) Let $A = 23$, $B = 5$. Find Q, R.
(b) Let $A = -23$, and $B = 5$. Find Q, R.
(c) For $A >= 0, B > 0$ show Q, R are unique.

Solution.
(a) $(Q, R) = (4, 3)$ since $23 = 5 \cdot 4 + 3$.
(b) $-23 = 5 \cdot (4) - 43$ but also $-23 = 5 \cdot (-4) - 3$ but also $-23 = 5 \cdot (-5) + 2$. In this latter case $(Q, R) = (-5, 2)$.All other possibilities are discarded.
(c) Let there exist two pairs Q_1, R_1 and Q_2, R_2 such that $A = BQ_1 + R_1$ and $A = BQ_2 + R_2$. Moreover $0 \leq R < B$ i.e. $0 \leq R_1 < B$ and $0 < R_2 < B$. Then $R_1 \neq R_2$ and we may assume $R_1 > R_2$ without loss of generality. Why?
If $R_1 = R_2$ then $A - R_1 = A - R_2$ and then $BQ_1 = BQ_2$. Since $B > 0$ we divide by B and then $Q_1 = Q_2$ in addition to $R_1 = R_2$. We proceed with the existence of two remainders and $R_1 > R_2$. Take $A = BQ_1 + R_1 = BQ_2 + R_2$ We have that $B(Q_2 - Q_1) = R_1 - R_2$. Because $0 \leq R_1, R_2 < B$ and B divides $B(Q_2 - Q_1)$ then B divides the equal $R_1 - R_2$. But the difference $R_1 - R_2 < B$. The only way for B to divide $R_1 - R_2$ is when $R_1 = R_2$. Done (In addition we have $Q_1 = Q_2$ as derived earlier for the other case.)
Equivalently from $B(Q_2 - Q_1) = R_1 - R_2$ we could argue that $Q_2 - Q_1 = (R_1 - R_2)/B$ is an integer. (The diffence of two integer Q_1, Q_2 is an integer.) Since $0 \leq R_1, R_2 < B$ then $0 \leq (R_1 - R_2)/B < 1$. The only integer that can satisfy this is $R_1 - R_2 = 0$. ■

Exercise 98.
What does this program do

```
x = x + y ;
y = y - x ;
x = x + y ;
y = -y    ;
```

Solution.
Value x gets swapped with value y! ■

Chapter 3

Sequences

3.1 Short review on sequences and their sums

3.1.1 Finite sequence

Definition 3.1 **Finite sequence**

A sequence of real numbers is finite if

$$(b_1, b_2, \ldots, b_n) \in \mathbb{R}^{\{1,2,\ldots,n\}}, \quad \text{or} \quad \{b_i\}_{i=1}^{n},$$

is a function $b : \mathbb{N} \to \mathbb{R}$, with $b(i) = b_i$, for all $i = 1, \ldots, n$. The numbers $b_1, b_2, \ldots, b_n$ are called the terms of the sequence.

3.1.2 Infinite sequence

Definition 3.2 **Ininite sequence**

A sequence of real numbers is infinite if

$$(a_1, a_2, \ldots, a_n, \ldots) \in \mathbb{R}^{\mathbb{N}}, \quad \text{or} \quad \{a_n\}_{n=1}^{\infty},$$

is a function $a : \mathbb{N} \to \mathbb{R}$, with $a(i) = a_i$, for all $i = 1, \ldots, n, \ldots$. The numbers $a_1, a_2, \ldots$ are called the terms of the sequence.

3.1.3 Summation and product of terms of a sequence

The summation symbol is used to add the terms of a sequence. If a sequence is finite such as

$$(b_1, b_2, \ldots, b_n) \in \mathbb{R}^{\{1,2,\ldots,n\}}.$$

we can define the sum of its terms as follows

$$\sum_{i=1}^{i=n} b_i = b_1 + \ldots + b_n.$$

Several times this gets written as follows.

$$\sum_{i=1}^{n} b_i = b_1 + \ldots + b_n,$$

or

$$\sum_i b_i = b_1 + \ldots + b_n,$$

with index values (upper or lower bound) implied.

A product can be defined similarly.

$$\prod_{i=1}^{i=n} = \prod_{i=1}^{n} = \prod_i = b_i = b_1 \cdot \ldots \cdot b_n.$$

For an infinite sequence, we similarly have the following.

$$\sum_{i=1}^{\infty} a_i = a_1 + \ldots + a_n + \ldots.$$

The product case follows similarly.

$$\prod_{i=1}^{\infty} a_i = a_1 \cdot \ldots \cdot a_n \cdot \ldots.$$

3.1.4 Series versus sum

The sums of the terms of sequence are ordinary finite. We formally refer to a sum as a series when the number of its terms is infinite. But we would also use the same name series for a finite sum.

3.1.5 Arithmetic sum

Definition 3.3 **Arithmetic sum**

The arithmetic sequence a_i, $i = 1, \ldots,$, is given by $a_i = i$ for all $i = 1, \ldots,$. We sum all the terms of the arithmetic sequence to n and call the sum $A(n)$ an arithmetic sum and as note before also an arithmetic series.

$$A(n) = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} i.$$

A closed-form expression for $A(n)$ is shown below.

$$A(n) = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Closed-form means or indicates the most compact and simplified form that can be assigned to a sum/series or product.

3.1.6 Recursive sequences

Definition 3.4 **Recursive function**

A recursive function is a function that invokes itself.

Definition 3.5 **Direct or indirect recursion**

In direct recursion a recursive function f invokes directly itself, whereas in indirect recursion, function f invokes function g that invokes f.

Definition 3.6 **Fibonacci sequence**

A well-known recursive function is the Fibonacci sequence (or function) defined as follows.

$$\{F_n\}_{n=0}^{\infty},$$

or it is sometimes described as

$$\{F_n\}_{n=1}^{\infty},$$

where for $n = 0, 1$ it is defined as follows,

$$F_0 = 0, F_1,$$

and for $n > 1$ it is defined as follows:

$$F_{n+1} = F_n + F_{n-1}.$$

The first few terms of the Fibonacci sequence are given as follows.

$$0, 1, 1, 2, 3, 5, \ldots$$

There is a mapping $F : \mathbb{N} \to \mathbb{N}$ that can be used to describe it, where $F(n)$ is equal to F_n. Thus

$$F_0, F_1, F_2, F_3, F_4, F_5, \ldots$$

maps to the sequence of integers (or natural numbers) indicated earlier.

3.1.7 Power series

Definition 3.7 **Power series with 0 center**

A power series $P(x)$ with center 0 is a series of the following form.

$$P(x) = \sum_{i=0}^{\infty} a_i x^i = a_0 + a_1 x + \ldots + a_i x^i + \ldots.$$

Definition 3.8 **Power series with c center**

A power series $P(x)$ with center c is a series of the following form.

$$P(x-c) = \sum_{i=0}^{\infty} a_i (x-c)^i.$$

For a power series we first consider the ratio of two consecutive terms r such that

$$r = \lim_{i \to \infty} |\frac{a_{i+1}(x-c)^{i+1}}{a_i (x-c)^i}|$$

That limit is equal to

$$r = |x-c||\lim \frac{a_{i+1}}{a_i}|$$

Let R be the reciprocal of $|\lim \frac{a_{i+1}}{a_i}|$. Then $r = |x-c|/R$. If $r < 1$ the series absolutely converges. If $r > 1$ it diverges. Equivalently $|x-c| < R$ and $|x-c| > R$ respectively. In other words the center of the interval of convergence, if it converges, is c and the radius is R.

Let $P(x-c) = \sum_{i=0}^{\infty} a_i (x-c)^i$ be a power series. There is an R such that $0 \leq R \leq \infty$ such that the series converges absolutely for $0 \leq |x-c| < R$ and diverges for $0 \leq |x-c| > R$. Furthermore, if $0 \leq r < R$, then the power series converges uniformly on the interval $|x-c| < r$ and the sum of the series is then continuous in $|x-c| < R$.

3.2 Infinite or finite sequences

Exercise 99.
Classify each sequence below as finite or infinite and increasing or decreasing.
(a) Sequence $\{1,2,3,5,7,11\}$.
(b) Sequence $a_n = 2^n$, $n \geq 150$.
(c) Sequence $a_n = 1/2^n$, $n \geq 1$.

Solution.
(a) It is finite and increasing.
(b) It is infinite and increasing.
(c) It is infinite and decreasing. ■

3.3 Fibonacci

Exercise 100.
Show the Euler-Binet formula that the n-th order Fibonacci number F_n, as defined earlier, is given by

$$F_n = \frac{1}{\sqrt{5}}\left[\phi^n - \left(-\frac{1}{\phi}\right)^n\right], \text{ where } \phi = \frac{1+\sqrt{5}}{2}.$$

Solution.
The linear recurrence $F_n = F_{n-1} + F_{n-2}$ can be rewritten $F_n - F_{n-1} - F_{n-2} = 0$. For $F_n = t^n$ one can show that t^n is a solution of the difference equation above if $t^2 - t - 1 = 0$. The latter has two solutions $t_1 = \phi$ and $t_2 = -1/\phi$. By linearity $F_n = at_1^n + bt_2^n$ is also a solution. Given that $F_0 = 0 = a + b$ and $F_1 = 1 = at_1 + bt_2$, we conclude $b = -a$ and $1 = at_1 + (-b)t_2$ which generates $a = 1/(\phi + 1/\phi)$. Substituting we conclude $a = 1/\sqrt{5} = -b$ and the result follows. ■

Exercise 101.

Let F_n be the n-th order Fibonacci number as defined earlier. Let $P(n)$ be the proposition shown below.

$$P(n) : \forall n \geq 0, F_n \leq 2^n.$$

Show that $P(n)$ is true for all $n \geq 1$.

Solution.
We prove the proposition by using strong induction for $n \geq 1$. The predicate $P(n)$ is as follows.

$$P(n) : F_n \leq 2^n.$$

STEP 1. Base case: Show that $P(0)$ is true.

P(0) is true is equivalent to showning $F_0 \leq 2^0$, which is true by definition since $F_0 = 0$. This completes the base case.

STEP 2. Inductive Step. $P(0) \wedge P(1) \wedge \ldots \wedge P(n-2) \wedge P(n-1) \Rightarrow P(n)$.

By the inductive (or induction) assumption (hypothesis) we first have that $P(n-2)$ is true i.e. $F_{n-2} \leq 2^{n-2}$, and we also have that $P(n-1)$ is true i.e. $F_{n-1} \leq 2^{n-1}$. Using the (recursive) definition for F_n we have the following.

$$\begin{aligned} F_n &= F_{n-1} + F_{n-2} \\ &= \underbrace{F_{n-1}}_{P(n-1)} + \underbrace{F_{n-2}}_{P(n-2)} \\ &\leq 2^{n-1} + 2^{n-2} \\ &\leq 2^{n-1} + 2^{n-1} \\ &\leq 2 \cdot 2^{n-1} \\ &\leq 2^n. \end{aligned}$$

$P(n)$ has been proven, concluding the proof of the inductive step and the inductive proof. ■

Exercise 102.

(a) If F_n is the order n fibonacci number, show the following equality.

$$F_1 + F_2 + F_3 + \ldots + F_n = F_{n+2} - 1.$$

(b) Show then the following: $f_1^2 + \ldots f_n^2 = f_n f_{n+1}$.

Solution.
(a) Use composition to generate the result.

$$\begin{aligned} F_1 &= F_3 - F_2 \\ F_2 &= F_4 - F_3 \\ F_3 &= F_5 - F_4 \\ &\ldots \\ F_{n-2} &= F_n - F_{n-1} \\ F_{n-1} &= F_{n+1} - F_n \\ F_n &= F_{n+2} - F_{n+1} \end{aligned}$$

Summing both side and eliminating major diagonals on the right hand-side we obtain the following.

$$F_1 + \ldots + F_n = F_{n+2} - F_n.$$

(b) Expand $f_{n-1} f_{n+1}$, and then generate multiple equalities for $n = 1, 2, \ldots$. ■

3.4 Binomial

Exercise 103.

Binomial theorem. For $a,b \in \mathbb{R}$ and $n \in \mathbb{N}$ we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k, \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Solution.
A proof follows by induction on n. We skip the proof of the base case for $n = 1$. It is obvious. Assuming (induction hypothesis) that the binomial theorem is true for n and thus

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

we are going to show that its form for $n+1$ is as follows.

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k,$$

We then form by utilizing the inductive assumption

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)^n \cdot (a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \cdot (a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^{k+1} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n+1-(k+1)} b^{k+1}. \qquad (3.1)
\end{aligned}$$

It suffices to show that the coefficient of $a^{n+1-k}b^k$ above in equality (3.1) is indeed $\binom{n+1}{k}$. From the first sum of equality (3.1) the coefficient of $a^{n+1-k}b^k$ is $\binom{n}{k}$. From the second sum of equality (3.1) the coefficient of $a^{n+1-k}b^k$ is $\binom{n}{k-1}$. Simple manipulations show that

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1},$$

a result that can be shown also through combinatorial interpretation. ■

3.5 Arithmetic, geometric and cubic sums

Exercise 104.

Arithmetic sum. The arithmetic sequence a_i, $i = 1, \ldots$, is given by $a_i = i$ for all $i = 1, \ldots$. We sum all the terms of the arithmetic sequence to n and denote the sum as $A(n)$. A closed-form expression for $A(n)$ is shown below.

$$A(n) = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Show that $A(n)$ is so, without using induction. (Every textbook in discrete mathematics available has an inductive proof of this result.)

Solution.
We write $A(n) = 1 + 2 + \ldots + (n-1) + n$ forwards as is and backwards and then we add up the corresponding terms.

$$\left.\begin{array}{rl} A(n) = & 1 + 2 + \ldots + (n-1) + n \\ A(n) = & n + (n-1) + \ldots + 2 + 1 \end{array}\right\} \text{Add up the two equations} \tag{3.2}$$

We have

$$\begin{array}{rcl} 2A(n) & = & (n+1) + (n+1) + \ldots + (n+1) + (n+1) \iff \\ 2A(n) & = & n(n+1) \iff \\ A(n) & = & n(n+1)/2. \end{array}$$

The expression $n(n+1)/2$ is the closed-form expression of the sum. It is compact and cannot be simplified further by making it more compact. ■

Exercise 105.

Arithmetic, quadratic and cubic sums. Let $A(n)$ be the arithmetic sum, let $Q(n)$ be the quadratic sum, and let $C(n)$ be the cubic sum. Their definitions are shown below along with a closed-form expression for $A(n), Q(n)$ and $C(n)$.

$$A(n) = \sum_{i=1}^{n} i = n(n+1)/2 \qquad Q(n) = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}, \qquad C(n) = \sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}.$$

Solution.
The sum for $A(n)$ was shown directly, without the use of induction, in the previous problem. We present below a method that derives the closed form for $A(n)$ through another method that can be generalized to obtain the closed form expressions for $Q(n)$ and $C(n)$.
We can use the following method to find sums of the following form

$$S_k = \sum_{i=1}^{n} i^k.$$

In this problem we resolve $A(n)$. First consider $(i+1)^{k+1}$ and expand it utilizing the binomial theorem. Substitute in the expansion $i = 1$, $i = 2, \ldots$, $i = n$, a total of n times and write the resulting n equalities one after the other. Then, add these n equalities by summing up the left hand sides and the right hand sides respectively. Solve for S_k and thus S_k can then be determined as a function of n.
For the sum $A(n)$ in question $k = 1$ and thus $S_1 = A(n)$. Therefore we consider

$$(i+1)^2 = i^2 + 2i + 1$$

We substitute for $i = 1, 2, \ldots, n$ writing one equality after the other

$$\begin{aligned}
(1+1)^2 &= 1^2 + 2\cdot 1 + 1 \\
(2+1)^2 &= 2^2 + 2\cdot 2 + 1 \\
(3+1)^2 &= 3^2 + 2\cdot 3 + 1 \\
(4+1)^2 &= 4^2 + 2\cdot 4 + 1 \\
\ldots &= \ldots \\
(n+1)^2 &= n^2 + 2\cdot n + 1
\end{aligned}$$

When we sum up the n equalities we realize that say, $(3+1)^2$ of the left-hand size of the third line is equal to 4^2 of the right-hand side of the fourth line and therefore the following can be derived after adding up these n equalities..

$$\begin{aligned}
(1+1)^2 + (2+1)^2 + \ldots + (n+1)^2 &= (1^2 + 2^2 + 3^2 + \ldots + n^2) + 2\cdot(1 + 2 + \ldots + n) + (1 + \ldots + 1) \\
(n+1)^2 &= (1^2) + 2\cdot(1 + 2 + \ldots + n) + (1 + \ldots + 1) \\
(n+1)^2 &= (1^2) + 2\cdot S_1 + (1 + \ldots + 1) \\
(n+1)^2 &= 1 + 2\cdot S_1 + n
\end{aligned}$$

We note that $2\cdot(1 + 2 + \ldots + n) = 2S_1$ and $(1 + \ldots + 1) = n$ (number of ones is number of equations). Then,

$$(n+1)^2 = 1 + 2S_1 + n$$

Solving for S_1 we get that $S_1 = ((n+1)^2 - n - 1)/2$, ie $S_1 = (n^2 + 2n + 1 - n - 1)/2 = (n^2 + n)/2 = n(n+1)/2$, which is $A(n) = S_1 = n(n+1)/2$. ■

Exercise 106.
Give a closed-form solution for the sum (i.e. compute the sum).

$$\sum_{i=0}^{n} \left(i^2 + i^3\right).$$

Solution.
The sum

$$\sum_{i=0}^{n} \left(i^2 + i^3\right),$$

can be rewritten as

$$\sum_{i=0}^{n} \left(i^2 + i^3\right) = \sum_{i=0}^{n} i^2 + \sum_{i=0}^{n} i^3.$$

Both the first and the second sum have well known closed form descriptions. For example, $\sum_{i=0}^{n} i^3 = n^2(n+1)^2/4$. Thus one way to derive the answer is some form of induction. We develop here a direct form. Let the first sum be denoted by S_2.

$$S_2 = \sum_{i=0}^{n} i^2$$

It is known that $S_1 = \sum_{i=0}^{n} i = n(n+1)/2$. This is because $S_1 = 1+2+\ldots+n$ and also $S_1 = n+(n-1)+\ldots+1$ and thus $2S_1 = (n+1)+(n+1)+\ldots+(n+1) = n(n+1)$ and the result follows. Moreover,

$$(i+1)^3 \quad = \quad i^3 + 3i^2 + 3i + 1.$$

We substitute $1, 2, 3, \ldots, n$ for i in the identity and we get a set of n equations

$$\begin{aligned}
(1+1)^3 &= 1^3 + 3\cdot 1^2 + 3\cdot 1 + 1 \\
(2+1)^3 &= 2^3 + 3\cdot 2^2 + 3\cdot 2 + 1 \\
(3+1)^3 &= 3^3 + 3\cdot 3^2 + 3\cdot 3 + 1 \\
\ldots &= \ldots \\
(n+1)^3 &= n^3 + 3\cdot n^2 + 3\cdot n + 1
\end{aligned}$$

By summing up these n equalities we obtain the following.

(Note that the left hand side of the i-th equality cancels out in the grand sum with the first term of the right hand side of the $(i+1)$-st equality, i.e. $(2+1)^3$ of the second equality cancels out in the sum with 3^3 of third equality).

$$\begin{aligned}
(n+1)^3 &= 1^3 + 3(\sum_{i=1}^{n} i^2) + 3(\sum_{i=1}^{n} i) + n \\
(n+1)^3 &= 1^3 + 3(\sum_{i=0}^{n} i) + 3(\sum_{i=0}^{n} i) + n \\
(n+1)^3 &= 1^3 + 3S_2 + 3S_1 + n \\
(n+1)^3 - 1^3 - n - 3S_1 &= 3S_2 \\
(n+1)^3 - (n+1) - 3n(n+1)/2 &= 3S_2 \\
(n+1)(n^2 + 2n + 1 - 1 - 3n/2) &= 3S_2 \\
(n+1)n(2n+1)/6 &= S_2
\end{aligned}$$

Therefore,

$$S_2 = \sum_{i=0}^{n} i^2 = n(n+1)(2n+1)/6.$$

S_3 can be computed similarly and so can even S_1. Then by adding up the expressions for S_2 and S_3 we get that

$$\sum_{i=0}^{n} \left(i^2 + i^3\right) = S_2 + S_3 = n(n+1)(2n+1)/6 + n^2(n+1)^2/4 = n(n+1)(3n^2+7n+2)/12.$$

■

3.6 Harmonic sequence and sum

Exercise 107.

Harmonic sum. Let $H(n)$ be the harmonic sequence sum. The definition of the harmonic sequence (series) is shown below.

$$H(n) = \sum_{i=1}^{n} \frac{1}{i} \approx \int (1/x)dx \approx \ln(n) + \gamma.$$

Show that $H(n)$ diverges as $n \to \infty$.

Solution.
We consider first the divergence of H_n through an example for H_8 that can be generalized for arbitrary n and not just 8. Consider H_8.

$$\begin{aligned} H_8 &= 1+1/2+1/3+1/4+1/5+1/6+1/7+1/8 \\ &\leq 1+1/2+1/2+1/4+1/4+1/4+1/4+1/8 \\ &\leq 1+1+1+1/8 = 3+1/8. \end{aligned}$$

Thus $H_8 \leq 3+1/8$. Likewise we can show then $H_{16} \leq 4+1/16$ and by induction $H_{2^k} \leq k+1/2^k$.
Going the other way around for a lower bound we have the following.

$$\begin{aligned} H_8 &= 1+1/2+1/3+1/4+1/5+1/6+1/7+1/8 \\ &\geq 1+1/2+1/4+1/4+1/8+1/8+1/8+1/8 \\ &\geq 1+1/2+1/2+1/2 \end{aligned}$$

By induction we can prove that $H_{2^k} \geq 1+k/2$. Moreover the following is true.

$$\int_1^{n+1} \frac{1}{x}dx < H_n < \int_1^n \frac{1}{x}dx + 1.$$

Then for $n > 1$,

$$\ln(n+1) < H_n < \ln(n) + 1.$$

Consider $\varepsilon_n = H_n - \ln(n)$. From the equation above we derive the following.

$$\begin{aligned} \ln(n+1) &< H_n < \ln(n)+1 \\ \ln(n+1)-\ln(n) &< H_n - \ln(n) < 1 \\ \ln(1+1/n) &< H_n - \ln(n) < 1 \\ \ln(1+1/n) &< \varepsilon_n < 1 \end{aligned}$$

Therefore

$$\varepsilon_n - \varepsilon_{n+1} = H_n - \ln(n) - H_{n+1} + \ln(n+1) = \ln(n+1) - \ln(n) - 1/(n+1).$$

Then

$$\varepsilon_n - \varepsilon_{n+1} = \int_n^{n+1} \frac{1}{x}dx - 1/(n+1).$$

Thus $H_n < \ln(n)+1$ can be strengthened to Thus $H_n \approx \ln(n)+\gamma$, where γ is Euler's constant. ■

Exercise 108.

For the harmonic series (sum of the terms of the harmonic sequence) show the following.

$$\sum_{i=1}^{n} H_i = (n+1)H_n - n,$$

$$\sum_{i=1}^{n} iH_i = \frac{n(n+1)}{2}H_n - \frac{n(n-1)}{4}.$$

Solution.
We provide a proof for the first expression. We note that

$$H_{n+1} = H_n + \frac{1}{n+1} \Rightarrow (n+1)H_{n+1} = (n+1)H_n + 1.$$

We expand the latter for $n+1, n, n-1, \ldots, 2$ as follows, after we move 1 to the other side

$$\begin{aligned}
(n+1)H_n &= (n+1)H_{n+1} - 1 \\
nH_{n-1} &= nH_n - 1 \\
(n-1)H_{n-2} &= (n-1)H_{n-1} - 1 \\
\ldots &= \ldots \\
2H_1 &= 2H_2 - 1
\end{aligned}$$

We then add the left-hand and right-hand sides, noting that there are n equations (shown and implied) altogether.

$$\begin{aligned}
H_n + H_{n-1} + \ldots + H_2 + 2H_1 &= (n+1)H_{n+1} - n \\
H_n + H_{n-1} + \ldots + H_2 + H_1 + 1 &= (n+1)H_{n+1} - n \\
\sum_{i=1}^{n} H_i + 1 &= (n+1)H_{n+1} - n
\end{aligned}$$

We then reuse $H_{n+1} = H_n + \frac{1}{n+1}$ to obtain the following.

$$\begin{aligned}
\sum_{i=1}^{n} H_i + 1 &= (n+1)H_{n+1} - n \\
\sum_{i=1}^{n} H_i + 1 &= (n+1)(H_n + \frac{1}{n+1}) - n \\
\sum_{i=1}^{n} H_i &= (n+1)H_n - n.
\end{aligned}$$

The result has been proven.

We now provide a proof for the second expression. The sum covers a triangular area.

$$\begin{aligned}
\sum_{i=1}^{n} iH_i &= \sum_{i=1}^{n} i \sum_{j=1}^{i} \frac{1}{j} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{i}{j} \\
&= \sum_{j=1}^{n} \frac{1}{j} \sum_{i=1}^{n} i \\
&= \sum_{j=1}^{n} \frac{1}{j} \left(\sum_{i=1}^{n} i - \sum_{i=1} j-1i \right) \\
&= \sum_{j=1}^{n} \frac{1}{j} \left(\frac{n(n+1)}{2} - \frac{j(j-1)}{2} \right) \\
&= \frac{n(n+1)}{2} \sum_{j=1}^{n} \frac{1}{j} - \frac{1}{2} \sum_{j=1}^{n} (j-1) \\
&= \frac{n(n+1)}{2} H_n - \frac{n(n-1)}{4}.
\end{aligned}$$

The same technique (sums) can be used to prove the first result as well. ■

3.7 Geometric sequence and sum

Exercise 109.

Geometric sum. Let $g_i = x^i$ for some $x \in \mathbb{R}$. Let for any natural number $n \geq 0$ and $x \in \mathbb{R}$ we define the geometric sum as follows.

$$G_n(x) = \sum_{i=0}^{n} x^i = x^0 + x^1 + \ldots + x^i + \ldots + x^n.$$

The sum is also known as the geometric series or as a sum of a finite number of terms of the geometric sequence. Show that for $n \geq 0$ and $x \in \mathbb{R}$ with $x \neq 1$, we have the following.

$$G_n(x) = \sum_{i=0}^{n} x^i = \frac{x^{n+1} - 1}{x - 1}.$$

Use a direct method to prove the result.

Solution.

The easiest way to show the result is by a direct method. We form the product $G_n(x)(x-1)$ by multiplying $\sum_{i=0}^{n} x^i = x^0 + x^1 + \ldots + x^i + \ldots + x^n$ by $x - 1$, and we show that the product is $x^{n+1} - 1$.

$$\begin{aligned} G_n(x)(x-1) &= \left(x^0 + x^1 + \ldots + x^{n-1} + x^n\right)(x-1) \\ &= \left(x^1 + x^2 + \ldots x^n + x^{n+1}\right) - \left(x^0 + x^1 + \ldots x^{n-1} + x^n\right) \\ &= \left(-x^0 + x^{n+1}\right) \\ &= \left(x^{n+1} - 1\right). \end{aligned}$$

Solving for $G_n(x)$ the result follows. ■

Exercise 110.

Geometric sum revisited.

Show that for $n \geq 0$ and $x \in \mathbb{R}$ with $x \neq 1$, we have the following.

$$G_n(x) = \sum_{i=0}^{n} x^i = \frac{x^{n+1}-1}{x-1}.$$

Use induction to prove the result.

Solution. We prove the theorem **by induction**.

We shall show that for $n \geq 1$ the following is true. (It is possible to show the result for $n \geq 0$ as well.) We call $G_n(x) = \frac{x^{n+1}-1}{x-1}$, with $x \neq 1$, the proposition $P(n)$.

$$G_n(x) = \sum_{i=0}^{n} x^i = \frac{x^{n+1}-1}{x-1}, \text{x} \neq 1.$$

Base case. $P(1)$ is true.

$$G_1(x) = 1 + x = \frac{x^2-1}{x-1}$$

which is obviously true for $x \neq 1$.

Inductive step. $P(n-1) \Rightarrow P(n)$. $P(n-1)$ is equivalent to

$$G_{n-1}(x) = \sum_{i=0}^{n-1} x^i = \frac{x^n-1}{x-1}, \text{x} \neq 1.$$

We have that $G_n(x) = G_{n-1}(x) + x^n$. We substitute the inductive step's right-hand side for $G_{n-1}(x)$ and we have the following derivations, for $x \neq 1$.

$$G_n(x) = G_{n-1}(x) + x^n = \frac{x^n-1}{x-1} + x^n = \frac{x^{n+1}-1}{x-1}$$

This establishes $P(n)$ from $P(n-1)$. ■

Exercise 111.

Sum of the terms of the geometric sequence becomes a series. We consider the geometric sequence $g_i = x_i$ for a range of $x \in \mathbb{R}$ so that $0 < x < 1$ or in general $|x| < 1$. Instead of forming a sum out of a finite number of terms of the sequence starting with $g_0, \ldots, g_n$ we form the series of all the terms of the sequence $g_0, \ldots, g_i, \ldots$. In other words we consider the case

$$G(x) = \lim_{n\to\infty} G_n(x).$$

$$G(x) = \sum_{i=0}^{\infty} x^i = x^0 + x^1 + \ldots + x^i + \ldots = \frac{1}{1-x}$$

Solution.
Consider $G_n(x)$ (one of the previous problems)

$$G_n(x) = \sum_{i=0}^{n} x^i = \frac{x^{n+1}-1}{x-1},$$

and take the limit of $G_n(x)$ for $n \to \infty$.

$$\begin{aligned} G(x) &= \lim_{n\to\infty} G_n(x) \\ &= \lim_{n\to\infty} \frac{x^{n+1}-1}{x-1} \\ &= \frac{0-1}{x-1} \\ &= \frac{1}{1-x}. \end{aligned}$$

Note that for $|x| < 1$ the $\lim_{n\to\infty} x^n = 0$. ■

Exercise 112.

Show that a closed form write-up of the sum

$$\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{\lg n-1}}$$

is $1-2/n$. This explains what **closed form** is.

Solution.
Let $X=\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{\lg n-1}}$. We use the fact that $1+x+\ldots+x^{m-1}=(x^m-1)/(x-1)$ for $x=1/2$ and $m=\lg n$.

$$\begin{aligned} 1+X &= 1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{\lg n-1}} \\ &= \frac{1/2^{\lg n}-1}{1/2-1} \\ &= \frac{1/n-1}{-1/2} \\ &= 2-2/n \end{aligned}$$

Since $1+X=2-2/n$ then $X=1-2/n$ and the result has been proved. ■

Exercise 113.

Geometric alternating sum. Let $g_i = (-x)^i$ for some $x \in \mathbb{R}$. Let for any natural number $n > 0$ and $x \in \mathbb{R}$ we define the alternating geometric sum as follows.

$$E_n(x) = \sum_{i=0}^{n} (-x)^i = x^0 - x^1 + \ldots + (-x)^i + \ldots + (-x)^n.$$

The sum is known as the geometric alternating sum of a finite number of the terms of the geometric alternating sequence.
Show that for $n > 0$ and $x \in \mathbb{R}$ with $x \neq -1$, we have the following.

$$G_n(x) = \sum_{i=0}^{n} (-x)^i = \frac{1 - (-x)^{n+1}}{1 + x}.$$

Note that for $x = -1$, $E_n(-1) = n + 1$. Use a direct method to prove the result.

Solution.
Start with the gemetric sum. Or use induction. ■

Exercise 114.

A sum that can be used for a BuildHeap proof. For $x \in \mathbb{R}$ and $x \neq 1$ we have the following finite sum.

$$I_n(x) = \sum_{i=0}^{n} ix^i = 0 \cdot x^0 + 1 \cdot x^1 + \ldots + i \cdot x^i + \ldots + n \cdot x^n.$$

Show that for $x \neq 1$, we have the following

$$I_n(x) = \sum_{i=0}^{n} ix^i = \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2}.$$

Solution.
By taking the first derivative with respect to x of $1 + x + x^2 + \ldots + x^n$ we obtain the derivations shown below.

$$\begin{aligned}
1 + x + x^2 + x^3 + \ldots + x^n &= \frac{x^{n+1} - 1}{x - 1} \\
&\quad (\textit{take first derivative of both sides}) \\
0 + 1 \cdot x^0 + 2 \cdot x^1 + 3 \cdot x^2 + \ldots + n \cdot x^{n-1} &= \frac{(n+1)x^n(x-1) - x^{n+1} + 1}{(x-1)^2} \\
0 + 1 \cdot x^0 + 2 \cdot x^1 + 3 \cdot x^2 + \ldots + n \cdot x^{n-1} &= \frac{nx^{n+1} - (n+1)x^n + 1}{(x-1)^2} \\
&\quad (\textit{mulitiply by x both sides}) \\
0 + 1 \cdot x^1 + 2 \cdot x^2 + 3 \cdot x^3 + \ldots + n \cdot x^n &= \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2} \\
x + 2x^2 + 3x^3 + \ldots + nx^n &= \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2} \\
\sum_{i=1}^{n} ix^i &= \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2}
\end{aligned}$$

Note 1. For $|x| < 1$ and $n \to \infty$ we have

$$x + 2x^2 + 3x^3 + \ldots + ix^i + \ldots = \sum_{i=0}^{\infty} ix^i = \frac{x}{(1-x)^2}.$$

Note 2. For $x = 2$ and n replaced by $\lg n$ we derive the following.

$$\begin{aligned}
\sum_{i=0}^{n} ix^i &= \frac{mx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2} = \\
\frac{\lg n x^{\lg n+2} - (\lg n + 1)x^{\lg n+1} + x}{(x-1)^2} &= \frac{\lg n 2^{\lg n+2} - (\lg n + 1)2^{\lg n+1} + 2}{(2-1)^2} \\
&= 4n \lg n - 2n(\lg n + 1) + 2 \\
&= 2n \lg n - 2n + 2.
\end{aligned}$$

Note 3. For $x = 1/2$ we have that $\sum_{i=0}^{n} ix^i = 2 - \frac{n+2}{2^n}$. ■

Exercise 115.

Series version of the previous problem's sum. Let for $x \in \mathbb{R}$ and $x \neq 1$ we define the following sum.

$$I_n(x) = \sum_{i=0}^{n} ix^i = 0 \cdot x^0 + 1 \cdot x^1 + \ldots + i \cdot x^i + \ldots + n \cdot x^n.$$

We have established previously the following.

$$I_n(x) = \sum_{i=0}^{n} ix^i = \frac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2}, \quad x \neq 1, x \in \mathbb{R}.$$

Let us define for $|x| < 1$ and $x \in \mathbb{R}$ the following.

$$I(x) = \lim_{n \to \infty} I_n(x).$$

Show that for $|x| < 1$ and $x \in \mathbb{R}$ the following.

$$I(x) = \sum_{i=0}^{\infty} ix^i = \frac{x}{(1-x)^2}.$$

Furthermore, show that

$$I(1/2) = 2.$$

Solution.
The limit $\lim_{n \to \infty} I_n(x)$, defined as $I(x)$, utilizing the right-hand side of $I_n(x)$ above for $|x| < 1$ is obviously $x/(1-x)^2$.
Substituting $1/2$ in the latter we obtain $I(1/2) = 2$.
We could have utilized the sum of the terms of the geometric sequence using the derivative transformation.
For any $|x| < 1$, we have.

$$\begin{aligned}
G(x) = \sum_{i=0}^{\infty} x^i &= \frac{1}{1-x} \Leftrightarrow \\
1 + x + \ldots + x^i + \ldots &= \frac{1}{1-x} \Leftrightarrow \\
\left(1 + x + \ldots + x^i + \ldots\right)' &= \left(\frac{1}{1-x}\right)' \Leftrightarrow \\
0 + 1 \cdot x^0 + \ldots + i \cdot x^{i-1} + \ldots &= \frac{1}{(1-x)^2} \Leftrightarrow \\
0 \cdot x + 1 \cdot x^1 + \ldots + i \cdot x^i + \ldots &= \frac{x}{(1-x)^2} \Leftrightarrow \\
I(x) = \sum_{i=0}^{\infty} i \cdot x^i &= \frac{x}{(1-x)^2}
\end{aligned}$$

From the last one for $x = 1/2$ we get $I(1/2) = 2$. ■

Exercise 116.
How much is the series

$$\sum_{i=0}^{\infty} i(i-1)x^i,$$

where $0 < x < 1$? Provide a closed-form answer. Explain the work.

Solution.
For the infinite geometric sequence for $0 < x < 1$, we have the ferived the following for the corresponding series (sum of the terms of the sequence).

$$\sum_{i=0}^{\infty} x^i = \frac{1}{(1-x)}$$

Taking the first derivative of both sides the following is derived.

$$\sum_{i=0}^{\infty} i \cdot x^{i-1} = \frac{1}{(1-x)^2}$$

Multiplying with x both sides we get the following.

$$\sum_{i=0}^{\infty} i \cdot x^i = \frac{x}{(1-x)^2}.$$

From the second derivation above we take the first derivative once more of both sides. We get

$$\sum_{i=0}^{\infty} i(i-1) \cdot x^{i-2} = \frac{2}{(1-x)^3}$$

We then multiply both sides with x^2. We get

$$\sum_{i=0}^{\infty} i(i-1) \cdot x^i = \frac{2x^2}{(1-x)^3}.$$

All of the steps require $0 < x < 1$. Result proven. ■

3.8 Series

Exercise 117.
Show that for $a,b \in \mathbb{R}$ we have the following.

$$ab \leq \frac{1}{2}(a^2+b^2).$$

Moreover, if in addition $a,b \geq 0$, we have the following.

$$\sqrt{ab} \leq \frac{1}{2}(a+b).$$

Solution.
This is equivalent to $(a-b)^2 \geq 0$, which is obviously true. Moreover, the second can be proven by substituting $\sqrt{a},\sqrt{b}$ for a,b respectively. The latter inequality is known as the geometric arithmetic mean and can be generalized for arbitrary $a_1,\ldots,a_n$ instead of a,b, provided $a_i \geq 0$ and $a_i \in \mathbb{R}$.

■

Exercise 118.

Show that the sequence $\{a_n\}_{n=1}^{\infty}$, where

$$a_n = (1+1/n)^n$$

is increasing (in fact strictly monotonically increasing) and converges to a limit between 2 and 3 known as e.

Solution.
It suffices to show that

$$a_{n+1} = \left(1+\frac{1}{n+1}\right)^{n+1} > \left(1+\frac{1}{n}\right)^n = a_n$$

Consider n times $1+1/n$ and once 1, thus form sequence b_i, $i=1,\ldots,n+1$, such $b_i = 1+1/n$ for $i=1,\ldots,n$ and $b_{n+1}=1$. By the arithmetic geometric mean theorem.

$$\begin{aligned}
\left(\frac{b_1+\ldots+b_{n+1}}{n+1}\right)^{n+1} &\geq b_1\cdot b_2\cdot\ldots\cdot b_{n+1} \Leftrightarrow \\
\left(\frac{n+2}{n+1}\right)^{n+1} &\geq \left(1+\frac{1}{n}\right)^n \Leftrightarrow \\
\left(1+\frac{1}{n+1}\right)^{n+1} &\geq \left(1+\frac{1}{n}\right)^n \Leftrightarrow \\
a_{n+1} &\geq a_n.
\end{aligned}$$

Note that all inequalities are in fact $>$ rather than $\leq$ as the values b_i chosen are not ALL equal to each other. Moreover, from the binomial theorem

$$\begin{aligned}
a_n = \left(1+\frac{1}{n}\right)^n &= \sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} \\
&= \left(1+\frac{n}{1}\frac{1}{n}+\frac{n(n-1)}{2!}\frac{1}{n^2}+\frac{n(n-1)(n-2)}{3!}\frac{1}{n^3}+\ldots\right) \\
&= \left(2+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\ldots\right)
\end{aligned}$$

Note that a_n can also be shown to be increasing since each one of the terms other than the constant 2 is increasing in n. Thus $a_n > 2$ for $n \geq 2$. Since $n! = n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1 \geq 2^{n-1}$, we have that

$$\begin{aligned}
a_n = \left(1+\frac{1}{n}\right)^n &= \left(2+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\ldots\right) \\
&\leq \left(2+\frac{1}{2}+\ldots+\frac{1}{2^{n-1}}\right) \\
&\leq \left(1+\frac{1}{2^0}+\frac{1}{2^1}+\ldots+\frac{1}{2^{n-1}}\right) \\
&\leq 3-\frac{1}{2^{n-1}} < 3.
\end{aligned}$$

The a_n converges to a limit between 2 and 3. ■

Exercise 119.
Show that

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Solution.
From the prior problem, we have

$$a_n = \left(1+\frac{1}{n}\right)^n \geq \left(2+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\ldots\right).$$

Fixing for $k < n$ we also have

$$a_n = \left(1+\frac{1}{n}\right)^n \geq \left(1+\frac{1}{1!}\left(1-\frac{0}{n}\right)+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\ldots+\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{(k-1)}{n}\right)\right).$$

From above for $n \to \infty$ with k fixed we have,

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \geq \sum_{i=0}^{k}\frac{1}{i!}.$$

With $k \to \infty$, we have

$$e \geq \sum_{i=0}^{\infty}\frac{1}{i!}$$

Moreover,

$$\begin{aligned} a_n = \left(1+\frac{1}{n}\right)^n &= \left(2+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\ldots\right) \\ &\leq \left(1+\frac{1}{1!}+\frac{1}{2!}+\ldots+\frac{1}{n!}\right) \end{aligned}$$

From above for $n \to \infty$ we have

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \leq \sum_{i=0}^{\infty}\frac{1}{i!}.$$

The result follows. ■

Exercise 120.
Show that

$$0 < e - \sum_{i=0}^{n} \frac{1}{i!} < \frac{1}{n \cdot n!}.$$

Solution.

$$\begin{aligned}
e - \sum_{i=0}^{n} \frac{1}{i!} &= \sum_{i=n+1}^{\infty} \frac{1}{i!} \\
&= \frac{1}{n!}\left(\frac{1}{n+1} + \frac{1}{(n+1)(n+2)} + \dots\right) \\
&\leq \frac{1}{n!}\left(\frac{1}{n+1} + \frac{1}{(n+1)^2} + \dots\right) \\
&\leq \frac{1}{n!}\left(-1 + 1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \dots\right) \\
&\leq \frac{1}{n!}\left(-1 + \frac{n+1}{n}\right) \\
&\leq \frac{1}{n \cdot n!}.
\end{aligned}$$

■

Exercise 121.

A power series $P(x)$ with center 0 is a series of the following form.

$$P(x) = \sum_{i=0}^{\infty} a_i x^i = a_0 + a_1 x + \ldots + a_i x^i + \ldots.$$

A power series $P(x)$ with center c is a series of the following form.

$$P(x-c) = \sum_{i=0}^{\infty} a_i (x-c)^i.$$

For a power series we first consider the ratio of two consecutive terms r such that

$$r = \lim_{i\to\infty} |\frac{a_{i+1}(x-c)^{i+1}}{a_i(x-c)^i}|$$

That limit is equal to

$$r = |x-c||\lim \frac{a_{i+1}}{a_i}|$$

Let R be the reciprocal of $|\lim \frac{a_{i+1}}{a_i}|$. Then $r = |x-c|/R$. If $r < 1$ the series absolutely converges. If $r > 1$ it diverges. Equivalently $|x-c| < R$ and $|x-c| > R$ respectively. In other words the center of the interval of convergence, if it converges, is c and the radius is R.
Let $P(x-c) = \sum_{i=0}^{\infty} a_i (x-c)^i$ be a power series. There is an R such that $0 \le R \le \infty$ such that the series converges absolutely for $0 \le |x-c| < R$ and diverges for $0 \le |x-c| > R$. Furthermore, if $0 \le r < R$, then the power series converges uniformly on the interval $|x-c| < r$ and the sum of the series is then continuous in $|x-c| < R$.

Prove the claims.

Solution.
Let w.l.o.g. $c = 0$; otherwise replace x with $x-c$. Say that $\sum_{i=0}^{\infty} a_i y^i$ converges for some $y \in R$ with $y \neq 0$. Its terms would then converge to zero and they would be bounded and there would exist a $B \ge 0$ such that $|a_n y^n| \le B$ for $n = 0, 1, \ldots$. If $|x| < |y|$ then

$$|a_n x^n| = |a_n y^n||\frac{x^n}{y^n}| \le Br^n,$$

where $r = |\frac{x}{y}| < 1$. Comparing $P(x)$ with $\sum Br^n$ we conclude that $P(x)$ is convergent; if the power series converges for some $y \in R$ then it converges absolutely for every $x \in R$ with $|x| < |y|$.
Let $R = \sum\{|x| \ge 0 : \sum a_i x^i\}$ converges. If $R = 0$ then the series converges for $x = 0$. If $R > 0$ then series converges absolutely for each $x \in R$ with $|x| < R$ because it converges for some $y \in R$ with $|x| < |y|$. By definition the series diverges for each $x \in R$ with $|x| > R$. If $R = \infty$ then the series converges for all $x \in R$. Let $0 < b < R$ and let $|x| \le b$. Choose a $c > 0$ such that $b < c < R$. Then $\sum |a_i c^i|$ converges, so $|a_i c^i| \le B$ and then

$$|a_i x^i| = |a_i c^i||\frac{x}{c}|^i \le |a_i c^i||\frac{b}{c}|^i \le Br^i$$

whewre $r = b/c < 1$. Since $\sum Br^i < \infty$ the series converge uniformly for $|x| \le b$ and the sum is continuous on $|x| \le b$. Given that this holds for each $0 \le b < R$, the sum is continuous in $|x| < R$. ■

Exercise 122.
A power series $P(x-c)$ with center c

$$P(x-c)=\sum_{i=0}^{\infty}a_i(x-c)^i$$

if it converges for $|x-c|<R$ and diverges for $|x-c|>R$ then $0\leq R\leq\infty$ is called the radious of convergence of the power series. Suppose that $a_i\neq 0$ for all sufficiently large i and the limit

$$R=\lim_{i\to\infty}|\frac{a_i}{a_{i+1}}|$$

exists or diverges to infinity. The $P(x-c)$ has radius of convergence R.

Solution.
Let

$$r=\lim_{i\to\infty}\frac{a_{i+1}(x-c)^{i+1}}{a_i(x-c)^i}|=|x-c|\lim|a_{i+1}/a_i.$$

The power series converges if $0\leq r<1$ or $|x-c|<R$ and diverges if $1<r\leq\infty$ or $|x-c|>R$. ■

Exercise 123.
The radius of convergence R of $P(x-c)$ is

$$R = \frac{1}{\limsup_{i\to\infty} |a_i|^{1/i}}$$

where $R = 0$ if the $\limsup$ diverges to ∞ and $R = \infty$ if $\lim\sum$ is 0.

Solution.
Let

$$r = \limsup |a_i(x-c)^i|^{1/i} = |x-c| \lim \sum a_i^{1/i}.$$

The series converges for $0 \leq r < 1$ or $|x-c| < R$ and diverges for $1 < r \leq \infty$ or $|x-c| > R$. ■

Exercise 124.
Examine the convergence of the series.

$$G(x) = \sum_{i=0}^{\infty} x^i$$

Solution.
The series

$$G(x) = \sum_{i=0}^{\infty} x^i$$

has radius of convergence $R = 1 = \frac{1}{\lim 1^{1/i}}$. It converge for $|x| < 1$ to $1/(1-x)$ and diverges for $|x| > 1$. At $x = 1$ or $x = -1$ the series diverges. The interval of convergence is $(-1, 1)$. ■

Exercise 125.
Examine the convergence of the series.

$$H(x) = \sum_{i=0}^{\infty} (1/i)x^i$$

Solution.
The series

$$H(x) = \sum_{i=0}^{\infty} (1/i)x^i$$

has radius of convergence $R = 1 = \lim 1/i/1/(i+1) = 1$. At $x = 1$ it is the harmonic series which diverges, at $x = -1$ it is the alternating harmoning series that converges but not absolutely. ■

Exercise 126.
Examine the convergence of the series.

$$K(x) = \sum_{i=0}^{\infty} (1/i!)x^i$$

Solution.

$$K(x) = \sum_{i=0}^{\infty} (1/i!)x^i$$

has radius of convergence $R = 1 = \lim 1/i!/1/(i+1)! = \infty$. It converges for all $x \in R$. It is $exp(x)$. ■

Exercise 127.

The following series is given. Examine the radius and interval of convergence.

$$L(x) = \sum_{i=0}^{\infty} (x+4)^i$$

Solution.
Apply the ratio test considering x a fixed value.

$$\lim_{i\to\infty} |\frac{(x+4)^{i+1}}{(x+4)^i}| = |x+4| < 1.$$

The radius of convergence is 1. The interval of convergence $|x+4| < 1$ implies $-1 < x+4 < 1$ implies $-5 < x < -3$. Thus in this interval it converges absolutely. We test the two endpoints -5 and -3. For $x = -5$ we have $L(-5) = \sum(-1)^i$ and the series diverges. For $L(-3) = \sum 1$ and it also diverges. Thus there is no conditional convergence. ■

3.9 Average seek time of a HDD

Exercise 128.

OS related. Assume a Hard-Disk Drive (HDD) contains $N+1$ tracks indexed 0 through N. The maximum seek time of an arm/heads movement, expressed in number of tracks, is N, when the heads move from track 0 to track N or the other way around. The average seek time, expressed in number of tracks, is

$$\frac{N}{3}+\frac{N}{3(N+1)}=\frac{N}{3}+\frac{1}{3}-\frac{1}{3(N+1)}$$

which is approximately,

$$\frac{N}{3}+\frac{1}{3}-\frac{1}{3(N+1)}\to\frac{N}{3}+\frac{1}{3}.$$

Show the result.

Solution.
If the arm/heads move from track i to track j, the distance in track covered is $|i-j|$. Thus the average seek time A, in terms of number of tracks, is the average over all initial and over all final positions of the arm/heads. The number of choices for i is $N+1$ (i.e. 0 through N) and likewise for j. Therefore Therefore

$$A = \frac{\sum_{i=0}^{N}\sum_{j=0}^{N}|i-j|}{(N+1)^2}=\frac{1}{(N+1)^2}\cdot\sum_{i=0}^{N}\sum_{j=0}^{N}|i-j|=\frac{1}{(N+1)^2}\cdot S. \tag{3.3}$$

We compute next the sum S.

$$\begin{aligned}
S &= \sum_{i=0}^{N}\sum_{j=0}^{N}|i-j| && (3.4)\\
&= \sum_{i=0}^{N}\left[\sum_{j=0}^{i}|i-j|+\sum_{j=i+1}^{N}|i-j|\right] && (3.5)\\
&= \sum_{i=0}^{N}\left[\sum_{j=0}^{i}(i-j)+\sum_{j=i+1}^{N}(j-i)\right] && (3.6)\\
&= \sum_{i=0}^{N}\left[\sum_{j=0}^{i}i-\sum_{j=0}^{i}j+\sum_{j=i+1}^{N}j-\sum_{j=i+1}^{N}i\right] && (3.7)\\
&= \sum_{i=0}^{N}\left[i(i+1)-i(i+1)/2+\sum_{j=0}^{N}j-\sum_{j=0}^{i}j-i(N-i)\right] && (3.8)\\
&= \sum_{i=0}^{N}[i(i+1)-i(i+1)/2+N(N+1)/2-i(i+1)/2-i(N-i)] && (3.9)\\
&= \sum_{i=0}^{N}[N(N+1)/2-i(N-i)] && (3.10)\\
&= \sum_{i=0}^{N}N(N+1)/2-N\cdot\sum_{i=0}^{N}i+\sum_{i=0}^{N}i^2 && (3.11)\\
&= N(N+1)^2/2-N^2(N+1)/2+N(N+1)(2N+1)/6 && (3.12)
\end{aligned}$$

After some minor calculations we obtain the following

$$\begin{aligned} S &= \sum_{i=0}^{N}\sum_{j=0}^{N}|i-j| && (3.13)\\ &= \frac{3N(N^2+2N+1)-3N^3-3N^2+2N^3+3N^2+N}{6} && (3.14)\\ &= \frac{3N^3+6N^2+3N-3N^3-3N^2+2N^3+3N^2+N}{6} && (3.15)\\ &= \frac{2N^3+6N^2+4N}{6} = \frac{N^3+3N^2+2N}{3} && (3.16)\\ &= N(N+1)(N+2)/3. && (3.17) \end{aligned}$$

Therefore from Equation 3.13 and its final dreivation, by replacing S into Equation 3.3 we obtain the following.

$$\begin{aligned} A &= \frac{1}{(N+1)^2}\cdot S = \frac{1}{(N+1)^2}\cdot\frac{N(N+1)(N+2)}{3} = \frac{N}{3}+\frac{N}{3(N+1)} = \frac{N}{3}+\frac{N+1-1}{3(N+1)} && (3.18)\\ &= \frac{N}{3}+\frac{1}{3}-\frac{1}{3(N+1)} && (3.19) \end{aligned}$$

■

Chapter 4

Identities and Inequalities

4.1 Short review on inequalities

All numbers in the remainder are real numbers, unless otherwise stated.

The product of two numbers is denoted by the binary operator that is nothing or by $\cdot$ or by $*$.

4.1.1 Quadratic polynomial

Let

$$f(x) = ax^2 + bx + c,\ a,b,d \in \mathbb{R}, \quad a \neq 0.$$

Polynomial f can be rewritten as follows.

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{\Delta}{4a}, \quad \Delta = b^2 - 4ac.$$

Theorem 4.1

Let $f(x)$ be as defined before. Let $a > 0$ and $x_0 = -b/(2a)$. Then the following apply.

1. *$f(x) > f(x_0)$ for any $x \neq x_0$.*
2. *$f(x) > 0$ for all $x \in \mathbb{R}$ if and only if $\Delta < 0$.*
3. *If $\Delta = 0$ then $f(x) \geq 0$. It is $f(x) = 0$ if and only if $x = x_0$.*
4. *If $\Delta > 0$ then $f(x)$ has two distinct roots $x_1 < x_2$. Then for $x < x_1$ or $x > x_2$, $f(x) > 0$ and $f(x) < 0$ for $x_1 < x < x_2$.*

For the case $a < 0$ symmetry applies.

Proof.

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{\Delta}{4a} \geq -\frac{\Delta}{4a} = f(x_0)$$

Equality above applies for $a\left(x + \frac{b}{2a}\right)^2 = 0$, i.e. $x = x_0$. This complete part (i). For $f(x) > 0$ we need $\Delta < 0$ to complete part (ii). Part (iii) draws from part (i) and part (ii). For part (iv) we write

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{\Delta}{4a} = a\left(x + \frac{b}{2a}\right)^2 - a\frac{\Delta}{4a^2} = a\left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{\Delta}}{2a}\right)^2\right)$$

We then obtain the following.

$$f(x) = a\left(\left(x - \left(\frac{-b-\sqrt{\Delta}}{2a}\right)\right)\left(x - \left(\frac{-b+\sqrt{\Delta}}{2a}\right)\right)\right) = a(x-x_1)(x-x_2).$$

■

4.1.2 Methods to show an inequality

Method 4.1 **Method of difference**

For $A, B \in \mathbb{R}$, in order to show $A > B$ we take the difference $A - B$ and show it is positive that is $A - B > 0$. For $A, B \in \mathbb{R}$, in order to show $A \geq B$ we take the difference $A - B$ and show it is non-negative that is $A - B \geq 0$.

If we want to show $A < B$ we prove $A - B$ is negative and other cases are handled similarly.

Method 4.2 **Method of division**

For $A, B \in \mathbb{R}$ and $A, B > 0$, in order to show $A > B$ we take the fraction A/B and show it is greater than one, that is $A/B > 1$. For $A, B \in \mathbb{R}$ and $A, B > 0$, in order to show $A \geq B$ we take the fraction A/B and show it is greater than or equal to one, that is $A/B \geq 1$.

Method 4.3 **Method of logarithms**

For $A, B \in \mathbb{R}$ and $A, B > 0$, in order to show $A > B$ we take the logarithms $\lg A$ (base two), or $\ln A$ (base e), and $\lg B$ (base two), or $\ln B$ (base e), and then compare with any one of these methods the same like logarithms e.g. $\lg A$ and $\lg B$.

Method 4.4 **Method of exponentials**

For $A, B \in \mathbb{R}$ and $A, B > 0$, in order to show $A > B$ we raise two or e to a power that is A or B, 2^A (base two), or e^A (base e), and 2^B (base two), or e^B (base e), and then compare with any one of these methods the same like exponentials e.g. e^A and e^B.

Method 4.5 **Method of powers**

For $A, B \in \mathbb{R}$ and $A, B > 0$, in order to show $A > B$ we raise these bases to the same power, for example A^2 and B^2 and then compare with any one of these methods the same like powers.

4.1.3 Convexity and concavity

Theorem 4.2

Let f be a function defined on (a,b) and let for any $x \in (a,b)$ there exists a second derivative $f''(x)$. The function f is convex if and only if for each $x \in (a,b)$ we have $f''(x) \geq 0$. The function f is strictly convex if and only if for each $x \in (a,b)$ we have $f''(x) > 0$. The function f is concave if and only if for each $x \in (a,b)$ we have $f''(x) \leq 0$. The function f is strictly concave if and only if for each $x \in (a,b)$ we have $f''(x) < 0$.

4.1.4 Function minima and maxima

Definition 4.1 **Maximum of a function**

A function $f(x)$ on domain X has a global/absolute maximum at $x = M$, $x \in d(f) = X$, if $f(x) \leq f(M)$ for every $x \in X$. $f(M)$ is the global/absolute maximum value of f in domain X.

Definition 4.2 **Minimum of a function**

A function $f(x)$ on domain X has a global/absolute minimum at $x = m$, $x \in d(f) = X$, if $f(x) \geq f(m)$ for every $x \in X$. $f(m)$ is the global/absolute minimum value of f in domain X.

The minimum and maximum values are known as the extreme values of f.

Definition 4.3 **Local maximum of a function**

A function $f(x)$ on domain X has a local maximum at $x = M$ if $f(x) \leq f(M)$ for every $x \in (c,d)$, where $(c,d) \subseteq X$ and $M \in (c,d)$. $f(M)$ is the local maximum value of f around M.

Definition 4.4 **Local minimum of a function**

A function $f(x)$ on domain X has a local minimum at $x = m$ if $f(x) \geq f(m)$ for every $x \in (c,d)$, where $(c,d) \subseteq X$ and $m \in (c,d)$. $f(m)$ is the local minimum value of f around m.

Theorem 4.3 **Extreme value**

If function f is real-valued with $d(f) = R$ and is continuous in the closed and bounded interval $[a,b]$ then f must attain an absolute/global minimum and an absolute/global maximum in $[a,b]$ at least once. Therefore there exist $m, M \in [a,b]$ such that

$$f(m) \leq f(x) \leq f(M), \quad \forall x \in [a,b].$$

We assume that the closed interval is bounded.

No local minimum can exist at an end-point a, b of a closed interval $[a,b]$. For all other points a global/absolute minimum/maximum is also a local minimum/maximum.

Thus per the extreme value theorem, f as in the theorem, has an absolute/global minimum and maximum on $[a,b]$ and these are either at a and b or at an intermediate point c and then f has a local maximum/local minimum.

Definition 4.5 **Bounded from above and supremum**

Let X be a set of real numbers. Set X is bounded from above if there exists some real number M such that $x \leq M$ for every $x \in X$. The number M is called an upper bound of X.

Note that M may not be in X.

Definition 4.6 **Bounded from below and infimum**

Let X be a set of real numbers. Set X is bounded from below if there exists some real number m such that $x \geq m$ for every $x \in X$. The number m is called a lower bound of X.

Note that m may not be in X.

Definition 4.7 **Infimum**

Let X be a partially ordered set and Y a subset of it. The infimum of S, denoted by $inf(S)$, is the largest element of X that is smaller than or equal to every element of Y.

Definition 4.8 **Supremum**

Let X be a partially ordered set and Y a subset of it. The supremum of S, denoted by $sup(S)$, is the smallest element of X that is greater than or equal to every element of Y.

Theorem 4.4

If f has a local maximum or minimum at c and if $f'(c)$ exists then $f'(c) = 0$.

Definition 4.9 **Critical point**

For a function f with domain $d(f)$ a critical point c is a point such that $f'(c) = 0$ (and thus exists) or $f'(c)$ does not exist.

Theorem 4.5

For a function f as defined, if f has a local miminum/local maximum at c then c is a critical point i.e. $f'(c) = 0$ or $f'(c)$ does not exist.

Method 4.6

For a function f that continuous and bounded in the interval $[a,b]$, in order to find the max or min value of f

(a) we find the critical points and the corresponding values in $[a,b]$,

(b) we find for a,b, values $f(a)$ and $f(b)$, and

(c) the largest value of (a) and (b) determines the max value of f, and the min the minimum value of f in $[a,b]$.

For f, f', f'' being continuous if there exists a c such that $f'(c) = 0$ (critical point), then $f''(c) > 0$ means $f(c)$ is a local minimum, and then $f''(c) < 0$ means $f(c)$ is a local maximum. For $f''(c) = 0$ we have no information. (Example: Consider $f(x) = x^3$.)

4.2 Simple inequalities

Exercise 129.
Show that for $0 < a < b$ we have that $a^2 < b^2$.

Solution.

$$\begin{aligned}
&\text{Use R3 with } p = a \wedge p = a > 0: && a < b \Rightarrow && a^2 < ab && (4.1)\\
&\text{Use R3 with } p = b \wedge p = b > 0: && a < b \Rightarrow && ab < b^2 && (4.2)\\
&\text{Use Transitivity and } (4.1) \text{ and } (4.2) \text{ above}: && a^2 < ab \wedge ab < b^2 \Rightarrow && a^2 < b^2 &&
\end{aligned}$$

■

Exercise 130.
Show that for $0 < a < b$ we have that $a^3 < b^3$.

Solution.
Similar to the previous one. ■

Exercise 131.
Show that for any $n > 1$, $n^2 - 1 > 0$.

Solution.
Omitted. Practice makes perfect. ■

Exercise 132.
Transitivity. The two functions $f(n) = 25n^2$ and $g(n) = 25n^2 + 10$ can be compared by starting say from $f(n)$ and using transitivity. Do so.

Solution.

$$f(n) = 25n^2 < 25n^2 + 1 < 25n^2 + 2 < \ldots < 25n^2 + 10 = g(n)$$

(Transitivity is implicitly used $A < B < C$ means $A < B$ and $B < C$ and thus by transitivity $A < C$.) ■

Exercise 133.
Difference. The two functions $f(n) = 25n^2$ and $g(n) = 25n^2 + 10$ can be compared by taking the difference. Show how to do so.

Solution.
$g(n) - f(n) = 25n^2 + 10 - 25n^2 = 10 > 0$ and observing that $g(n) - f(n)$ is positive i.e. $g(n) > f(n)$. Equivalently $f(n) - g(n) < 0$ and also $f(n) < g(n)$. ■

Exercise 134.
Ratio. The two functions $f(n) = 25n^2$ and $g(n) = 25n^2 + 10$ can be compared by taking their ratio Take the ratio to compare the two functions. Do so.

Solution.
$g(n)/f(n) = (25n^2 + 10)/25n^2 \geq 1$ and observing that $g(n)/f(n) > 1$ i.e. $g(n) > f(n)$. Equivalently $f(n)/g(n) < 1$ and also $f(n) < g(n)$. ■

Exercise 135.
Exponentiation. Let $a(n) = \ln(25n^2)$ and $b(n) = \ln(25n^2 + 10)$. Raise the two functions on a common base and then compare. Do so.

Solution.
Then $f(n) = e^{a(n)} = 25n^2$ and $g(n) = e^{b(n)} = 25n^2 + 10$ The resulting functions $f(n)$ and $g(n)$ have already been compared and found to be such that $f(n) < g(n)$. Thus $a(n) < b(n)$ as well by monotonicity! ■

Exercise 136.
Logarithms. The two functions $f(n) = 25n^2$ and $g(n) = 25n^2 + 10$ can be compared by taking their logarithms. Take the logarithms and then compare. Do so.

Solution.
$a(n) = \ln(25n^2)$ and $b(n) = \ln(25n^2 + 10)$. We just proved above that $a(n) < b(n)$ and thus $f(n) < g(n)$ by monotonicity. (Admittedly, not a very interesting example!) ■

Exercise 137.
Compare $n^{1/\lg n}$ and 2.

Solution.
They are equal to each other.

$$\begin{aligned} n^{1/\lg n} &: 2 \\ \lg(n^{1/\lg n}) &: \lg 2 \\ (1/\lg n)\cdot \lg n &: 1 \\ 1 &: 1. \end{aligned}$$

Obviously $1 = 1$ and this $n^{1/\lg n}$ is equal to 2. ■

Exercise 138.
Let $1 \leq i \leq n$. Then $i(n+1-i) \geq n$.

Solution.
Since $n \geq i$ and $i-1 \geq 0$ multiplying with $i-1$ both sides of the former we get. get

$$n(i-1) \geq i(i-1) \Rightarrow n+n(i-1) \geq i(i-1)+n \Rightarrow n+n(i-1)-i(i-1) \geq i(i-1)+n-i(i-1) \Rightarrow i(n-i+1) \geq n.$$

■

4.3 Absolute value inequalities

Exercise 139.
(a) Show that for $a,b \in \mathbb{R}$ the following is true

$$|a|+|b| \geq |a+b|.$$

(b) Show then the following.

$$|a-b| \geq ||a|-|b||.$$

In the two inequalities above replacing b with $-b$ one then derives the following.

$$|a|+|b| \geq |a-b|,$$

and

$$|a+b| \geq ||a|-|b||.$$

Therefore we have

$$|a|+|b| \geq |a+b| \geq ||a|-|b||,$$

and

$$|a|+|b| \geq |a-b| \geq ||a|-|b||.$$

Solution.
(a) Let $|a| = \sqrt{a^2}$, let $|b| = \sqrt{b^2}$, let $|a+b| = \sqrt{(a+b)^2}$. Then it suffices to prove the following.

$$\begin{aligned} \sqrt{a^2}+\sqrt{b^2} &\geq \sqrt{(a+b)^2} \\ (\sqrt{a^2}+\sqrt{b^2})^2 &\geq (\sqrt{(a+b)^2})^2 \\ a^2+b^2+2\sqrt{a^2b^2} &\geq a^2+b^2+2ab \\ \sqrt{a^2b^2} &\geq ab \end{aligned}$$

Given that $\sqrt{a^2b^2} = \sqrt{(ab)^2} = |ab|$,

$$\begin{aligned} \sqrt{a^2b^2} &\geq ab \Leftrightarrow \\ |ab| &\geq ab, \end{aligned}$$

which is true for all a,b, and also $z = ab$.
(a) A direct way to show the previous result is as follows.

$$(|x|+|y|)^2 = |x|^2+|y|^2+2|x||y| = x^2+y^2+2|x||y| \geq x^2+y^2+2xy = (x+y)^2 = |x+y|^2.$$

From the latter we obtain $|x+y| \leq |x|+|y|$ as all quantities are non-negative.
(a) Another way to prove this is as follows

$$-|a| \leq a \leq |a|, \quad -|b| \leq b \leq |b|$$

If we add those two inequalities we obtain the following.

$$-|a|-|b| \leq a+b \leq |a|+|b| \Leftrightarrow -(|a|+|b|) \leq a+b \leq |a|+|b|$$

The latter indicates the following

$$|a+b| \leq |a|+|b|$$

(b)

$$|a-b|+|b| \geq |a-b+b| \Leftrightarrow |a-b|+|b| \geq |a| \Leftrightarrow |a-b| \geq |a|-|b|.$$

Similarly,

$$|b-a|+|a| \geq |b-a+a| \Leftrightarrow |b-a|+|a| \geq |b| \Leftrightarrow |b-a| \geq |b|-|a| \Leftrightarrow |a-b| \geq |b|-|a|.$$

We have that

$$||a|-|b|| = max\{|a|-|b|, -(|a|-|b|)\} = max\{|a|-|b|, |b|-|a|)\}$$

The latter along with the previous two inequalities implies

$$|a-b| \geq max\{|a|-|b|, |b|-|a|)\} = ||a|-|b||.$$

■

4.4 Factorial and combinatorials

Exercise 140.
Show that for every integer $n \geq 1$, we have that

$$n! \leq n^n.$$

Let $\lg n$ be the base two logarithm of n i.e. $2^{\lg n} = n$.
Then conclude that $\lg(n!) \leq n \lg n$ for all $n \geq 1$.

Solution.
The first result is a straighforward upper bound of each term of the factorial with n. Starting with $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$, we have that

$$\begin{aligned} n! &= \overbrace{1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n}^{n \text{ terms}} \\ &\leq \overbrace{n \cdot n \cdot \ldots \cdot n \cdot n}^{n \text{ terms}} \\ &\leq n^n \end{aligned}$$

If $n! \leq n^n$ by taking logarithms $\lg(n!) \leq n \lg n$. ■

Exercise 141.

For every $n \geq 1$, we have that

$$n! \geq (n/2)^{(n/2)}.$$

We also have the following.

$$n! \leq \left(\frac{n}{2}\right)^{\frac{n}{2}} n^{n/2} \quad \text{and} \quad n! \geq n^{\frac{n}{2}-\frac{\sqrt{n}}{2}}.$$

Solution.
The smallest half of the terms of the factorial are at least one (lower bound). The largest half of the terms of the factorial are at least $n/2$. Moreover $\lceil n/2 \rceil + \lfloor n/2 \rfloor = n$ and $\lfloor n \rfloor \leq n \leq \lceil n \rceil$. Therefore,

$$\begin{aligned}
n! &= \overbrace{1 \cdot 2 \cdot \ldots \cdot \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil \cdot \ldots \cdot (n-1) \cdot n}^{n \text{ terms}} \\
&\geq \overbrace{1 \cdot 1 \cdot \ldots \cdot 1 \cdot \lceil n/2 \rceil \cdot \ldots \cdot \lceil n/2 \rceil \cdot \lceil n/2 \rceil}^{n \text{ terms}} \\
&\geq \overbrace{1 \ldots 1}^{\lfloor n/2 \rfloor \text{ terms}} \overbrace{\lceil n/2 \rceil \ldots \lceil n/2 \rceil}^{\lceil n/2 \rceil \text{ terms}} \\
&\geq (\lceil n/2 \rceil)^{\lceil n/2 \rceil} \\
&\geq (n/2)^{(n/2)}.
\end{aligned}$$

If $n! \geq (n/2)^{(n/2)}$ by taking logarithms $\lg(n!) \geq (n/2)\lg(n/2)$. We have $\lg(n!) \geq (n/2)\lg(n/2)$ and $\lg(n!) \leq n \lg n$ i.e. $\lg(n!) \approx n \lg n$ utilizing the previous problem. A, by an example proof of the same, follows. For the floor or ceiling manipulations consider $n = 6$ with $n/2 = 3$ and $\lceil n/3 \rceil = \lfloor n/3 \rfloor = 2$.

$$6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \geq 1 \cdot 1 \cdot 1 \cdot 3 \cdot 3 \cdot 3 = 3^3 = (6/2)^{(6/2)}$$

Likewise, for $n = 5$ with $n/2 = 2.5$ and $\lfloor 5/2 \rfloor = 2$, $\lceil 5/2 \rceil = 3$.

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \geq 1 \cdot 1 \cdot 3 \cdot 3 \cdot 3 = 3^3 \geq (5/2)^{(5/2)}$$

The Stirling's approximated formula for $\mathbf{n}!$ states that for all $n > 10$, we have that $n! \approx \sqrt{2\pi n}(n/e)^n$ which is more formally defined as in

$$(n/e)^n \sqrt{2\pi n} e^{1/(12n+1)} \leq n! \leq (n/e)^n \sqrt{2\pi n} e^{1/(12n)}.$$

A simplified form of it is for $n > 10$ we have that $n! \approx \left(\frac{n}{e}\right)^n$.
Ignoring $\sqrt{n}$ terms we have the following.

$$\begin{aligned}
n! &= 1 \cdot 2 \cdot \ldots \cdot \sqrt{n} \cdot \ldots (n-1) \cdot n \\
&\geq 1 \cdot 1 \cdot \ldots \cdot \sqrt{n} \cdot \ldots (n-1) \cdot n \\
&\geq \left(\sqrt{n}\right)^{n-\sqrt{n}} \\
&\geq n^{\frac{n}{2}-\frac{\sqrt{n}}{2}}.
\end{aligned}$$

The corresponding upper bound can be proven similarly.

■

Exercise 142.

Stirling's approximation for the factorial. For any integer $n > 0$ we have the following.

$$(n/e)^n \leq n! \leq en(n/e)^n.$$

Solution.
See next problem or instead of using Stirling's approximation for the factorial we could use the following Taylor expansion for e^n.

$$e^n = \sum_{i=0}^{\infty} \frac{n^i}{i!} \geq \sum_{i=n}^{i=n} \frac{n^i}{i!} \geq \frac{n^n}{n!}$$

Thus $e^n \geq \frac{n^n}{n!}$ that yields the left part of the double inequality above. The other part involves a bit of calculus.

$$\ln(n!) = \sum_{i=1}^{n} \ln(k) \leq \int_1^n \ln(x)dx + \ln(n) \leq (x\ln x - x)|_1^n + \ln(n)$$

Expanding we have

$$\ln(n!) \leq (x\ln x - x)|_1^n + \ln(n) \leq (n\lg(n) - n + 1 + \ln(n).$$

And exponentiating this leads to the following conclusion.

$$n! \leq (n^n/e^n)en = en(n/e)^n,$$

as needed. ■

Exercise 143.
For any integer $n > 0$ and $1 \leq k \leq n$ we have the following.

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k. \tag{4.3}$$

Furthermore we will show later that for any integer $n > 00$ and $1 \leq k \leq n$, we have the following

$$\binom{n}{k} \leq 2^n,$$

using the binomial theorem.

Solution.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} \\
&= \frac{n^k}{k!} \quad \text{Stirling's approximation} \\
&\leq \frac{n^k}{(k/e)^k} \\
&= \left(\frac{en}{k}\right)^k
\end{aligned}$$

Instead of using Stirling's approximation for the factorial we could have used the following Taylor expansion for e^k.

$$e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} \geq \sum_{i=k}^{i=k} \frac{k^i}{i!} \geq \frac{k^k}{k!}$$

The last one implies $e^k/k^k \geq 1/k!$.
For the lower bound we have the following.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1} = \prod_{t=0}^{k-1} \frac{n-t}{k-t} \geq \prod_{t=0}^{k-1} \frac{n}{k} \geq \left(\frac{n}{k}\right)^k
\end{aligned}$$

For $t = 0$, $(n-t)/(k-t) = n/k \geq n/k$. For $t = 1$ through $t = k-1$, we have that $(n-t)/(k-t) \geq n/k$, as long as $k \leq n$. ■

Exercise 144.
The following holds.

$$\binom{2n}{n} = \frac{(2n)!}{n! \cdot n!} \geq 2^n.$$

Solution.

$$\begin{aligned}
\frac{(2n)!}{n! \cdot n!} &= \frac{(2n) \cdot (2n-1) \cdot \ldots \cdot (n+1)}{n \cdot (n-1) \cdot \ldots \cdot 1} \\
&= \frac{(2n) \cdot (2n-1) \cdot \ldots \cdot (2n-i) \cdot \ldots \cdot (2n-(n-1))}{n \cdot (n-1) \cdot \ldots \cdot (n-i) \cdot \ldots \cdot (n-(n-1))} \\
&= \frac{(2n)}{n} \cdot \frac{(2n-1)}{n-1} \cdot \ldots \cdot \frac{(2n-i)}{n-i} \cdot \ldots \cdot \frac{2n-(n-1)}{n-(n-1)} \\
&\geq 2 \cdot \frac{n}{n} \cdot 2 \cdot \frac{(n-1/2)}{n-1} \cdot \ldots \cdot 2 \cdot \frac{(n-i/2)}{n-i} \cdot \ldots \cdot 2 \cdot \frac{n-(n-1)/2}{n-(n-1)} \\
&\geq 2^n
\end{aligned}$$

The latter is due to the fact that

$$\frac{n-i/2}{n-i} \geq 1 \Leftrightarrow n - i/2 \geq n - i \Leftrightarrow i/2 \leq i.$$

■

Exercise 145.
(a)

$$\frac{2^{2n}}{2n} \leq \binom{2n}{n}. \tag{4.4}$$

(b) For any integer $n > 0$ we have the following.

$$2^n \leq \binom{2n}{n}. \tag{4.5}$$

(c) For any integer $n > 0$ we have the following.

$$\frac{2^{2n}}{2\sqrt{n}} \leq \binom{2n}{n} \leq \frac{2^{2n}}{\sqrt{2n}}. \tag{4.6}$$

Solution.
(a) From

$$(1+1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} \leq 1 + (2n-1)\binom{2n}{n} + 1 \leq 2n\binom{2n}{n},$$

utilizing the identity $2 \leq \binom{2n}{n}$. The two ones are for $k = 0$ and $k = 2n$.
(b) This is the previous problem. We use a different approach. Consider

$$2 \cdot 4 \cdot \ldots \cdot 2n \leq (n+1) \cdot (n+2) \cdot \ldots \cdot 2n,$$

by way of $2i \leq n+i$ for $i = 1, \ldots, n$. Then

$$\begin{aligned} 2 \cdot 4 \cdot \ldots \cdot 2n &\leq (n+1) \cdot (n+2) \cdot \ldots \cdot 2n \\ 2^n \cdot n! &\leq (n+1) \cdot (n+2) \cdot \ldots \cdot 2n \\ 2^n \cdot n! &\leq \frac{(2n)!}{n!} \\ 2^n &\leq \frac{(2n)!}{n! \cdot n!} \\ 2^n &\leq \binom{2n}{n} \end{aligned}$$

(c) Use Stirling's approximation for the factorial,

$$n! = \sqrt{2\pi n}(n/e)^n e^{a_n},$$

where $1/(12n+1) < a_n < 1/(12n)$. It can also be derived through Wallis's formula which is equivalent to

$$\lim_{n\to\infty} = \frac{(2n)!\sqrt{\pi n}}{(n!)^2 2^{2n}} = 1$$

Omitted. ■

4.5 Combinatorial identity bounds

Exercise 146.

Binomial theorem (identity). For $a,b \in \mathbb{R}$ and $n \in \mathbb{N}$ we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = a^n + na^{n-1}b + \frac{n(n-1)}{2}a^{n-2}b^2 + \ldots + \binom{n}{k}a^k b^{n-k} + \ldots + nab^{n-1} + b^n,$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Solution.
(a) A proof follows by induction on n. We skip the proof of the base case for $n = 1$. It is obvious. Assuming (induction hypothesis) that the binomial theorem is true for n and thus

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k,$$

we are going to show that its form for $n+1$ is as follows.

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k}b^k,$$

We then form by utilizing the inductive assumption

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)^n \cdot (a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k \cdot (a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k}b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^{k+1} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k}b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n+1-(k+1)}b^{k+1}. \qquad (4.7)
\end{aligned}$$

It suffices to show that the coefficient of $a^{n+1-k}b^k$ above in equality (4.7) is indeed $\binom{n+1}{k}$. From the first sum of equality (4.7) the coefficient of $a^{n+1-k}b^k$ is $\binom{n}{k}$. From the second sum of equality (4.7) the coefficient of $a^{n+1-k}b^k$ is $\binom{n}{k-1}$. Simple manipulations show that (a proof is deferred to the end of this discussion)

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1},$$

a result that can be shown also through combinatorial interpretation.
(b) The alternating sum case provides additional information on the various steps. For n we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = a^n + na^{n-1}b + \frac{n(n-1)}{2}a^{n-2}b^2 + \ldots + \binom{n}{k}a^k b^{n-k} + \ldots + nab^{n-1} + b^n.$$

Thus for $n+1$ we have

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)\cdot(a+b)^n \\
&= (a+b)\cdot\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} \\
&= (a+b)\cdot\left(a^n+na^{n-1}b+\frac{n(n-1)}{2}a^{n-2}b^2+\ldots+\binom{n}{k}a^k b^{n-k}+\ldots+nab^{n-1}+b^n\right) \\
&= a^{n+1-0}+\left(\binom{n}{1}+\binom{n}{0}\right)a^{n+1-1}b+\left(\binom{n}{2}+\binom{n}{1}\right)a^{n+1-2}b^2+\ldots+\left(\binom{n}{k}+\binom{n}{k-1}\right)a^{n+1-k}b^k+b^{n+1} \\
&= \sum_{k=0}^{n+1}\binom{n+1}{k}a^{n+1-k}b^k.
\end{aligned}$$

We use the equality

$$\binom{n}{k}+\binom{n}{k-1}=\binom{n+1}{k}$$

in the last expression. This can be proven directly from the left hand side to the right one, as noted earlier.

$$\begin{aligned}
\binom{n}{k}+\binom{n}{k-1} &= \frac{n!}{k!(n-k)!}+\frac{n!}{(k-1)!(n+1-k)!} \\
&= \frac{n!}{k!(n+1-k)!}\left((n+1-k)+k\right) \\
&= \frac{(n+1)\cdot n!}{k!(n+1-k)!} \\
&= \frac{(n+1)!}{k!(n+1-k)!} \\
&= \binom{n+1}{k}.
\end{aligned}$$

■

Exercise 147.
(a) For any n,k show that we have the following.

$$\binom{n}{k} = \binom{n}{n-k}$$

(b) Using the definition,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

show that

$$\binom{n}{k} = \binom{n}{n-k},$$

Solution.
(a) The number of ways to select k objects out of n is the number of ways to UNselect $n-k$ objects out of n thus selecting the remaining k objects.
(b) An easier proof than the one shown before. Rename k into $n-k$. ■

Exercise 148.
For any n,k show that we have the following.

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

This is the basis used in the Pascal Triangle to determine the coefficient of $a^k b^{n-k}$ i.e. $C(n,k)$ from the coefficients of $C(n-1,k)$ and $C(n-1,k-1)$. Note $C(n,k) = \binom{n}{k}$.

Solution. The number of ways to select k objects out of n is equal to the number of selections of k objects that contain 1 plus the number of selections of k objects that DO NOT contain 1. If the selection of k objects contains 1 the number of such selections is to select out of the (out of 1) remaining $n-1$ objects $k-1$ of them and add to the mix 1. This is $\binom{n-1}{k-1}$. If the selection of k objects does not contain 1 the number of such selections is to select out of the (out of 1) remaining $n-1$ objects k of them. This is $\binom{n-1}{k}$. ■

Exercise 149.
Using the definition

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k} = \frac{(-n)(-n-1)\dots(-n-k+1)}{k!}$$

show that

$$\binom{-1/2}{k} = \frac{(-1)^k}{4^k}\binom{2k}{k}$$

Solution.
Manipulate the latter. ■

Exercise 150.
(a) For all $n \in \mathbb{N}$ and a, b we have

$$(a^n - b^n) = (a-b)\left(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \ldots + ab^{n-2} + b\right).$$

Show that it is true by induction.
(b) For all $n \in \mathbb{N}$ and a we have

$$(a^n - 1) = (a-1)\left(a^{n-1} + a^{n-2} + a^{n-3} + \ldots + a + 1\right).$$

(c) For the previous theorem to the limit $n \to \infty$, and $-1 < a < 1$, we have $a^n \to 0$ and thus

$$1 + a + a^2 + \ldots + a^{n-1} = \frac{-1}{a-1} = \frac{1}{1-a}.$$

Solution.
(a) Direct manipulation. Omitted.
(b) In (a), set $b = 1$. It can be rewritten,

$$1 + a + a^2 + \ldots + a^{n-1} = \frac{a^n - 1}{a-1}.$$

as long as $a \neq 1$.
(c) In (b) use the limit $n \to \infty a^n = 1$, for $|a| < 1$. ■

Exercise 151.
Use the binomial theorem below to prove the following.
(a) Show that

$$(1+1)^n = 2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

(b) Show that

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k.$$

(c) For an alternative proof of (b) show that the number of subsets of $\{1,2,\ldots,n\}$ of even cardinality is equal to the number of subsets of $\{1,2,\ldots,n\}$ of odd cardinality and thus we have the following.

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k.$$

Solution.
(a) Set $a = b = 1$.
(b) Set $a = -b = 1$.

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k.$$

implies

$$\binom{n}{0} + \binom{n}{2} + \ldots = \binom{n}{1} + \binom{n}{3} + \ldots$$

■

Exercise 152.
For any integer $n > 0$ and $k \leq n$ we have the following.

$$\binom{n}{k} \leq 2^n.$$

Solution.
From the binomial theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^n b^{n-k},$$

set $a = b = 1$, and the right hand side sum is at least one of $n+1$ terms: one such term is obviously $\binom{n}{k}$. ■

Exercise 153.
It has been proven eearlier that the following holds.

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

The latter is derived for $a = b = 1$ above.

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n,$$

Show then that for $a = -1$ and $b = -x$ and $n = -n$ we have

$$(-1-x)^{-n} = \sum_{k=0}^{n} \binom{-n}{k} (-1)^k (-x)^{-n-k}$$

Solution.

$$\begin{aligned} (-1-x)^{-n} &= \sum_{k=0}^{-n} \binom{-n}{k} (-1)^{-n-k} (-x)^k \\ &= \sum_{k=0}^{n} \binom{-n}{k} (-1)^k (-x)^{-n-k} \end{aligned}$$

■

Exercise 154.
For any n,k,r, with $k \leq r \leq n$, we have the following.

$$\binom{n}{r}\binom{r}{k} = \binom{n}{k} + \binom{n-k}{r-k}$$

Solution.
The number of ways to choose k balls for bin RED and $r-k$ balls for bin BLUE is to select first r balls out of n and then out of the r balls choose the k balls to go to bin RED with the unchosen going to bin BLUE. This is the first left-hand side. Equivalently, First pick out of n the k RED bin balls. From the remaining $n-k$ ones pick the $r-k$ BLUE bin balls. ■

Exercise 155.
For any non-negative integer a, b, and r we have the following.

$$\sum_{k=0}^{r} \binom{a}{k}\binom{b}{r-k} = \binom{a+b}{r}$$

Solution.
We have $a+b$ call it n balls, a are RED and b are BLUE. In how many ways can we pick r balls out of n? ■

Exercise 156.
(a) Show for $k \leq (n-1)/2$ we have

$$\binom{n}{k} \leq \binom{n}{k+1}.$$

(b) Show for $k \geq (n-1)/2$ we have

$$\binom{n}{k} \geq \binom{n}{k+1}.$$

(c) The largest of $\binom{n}{i}$, $i = 0, \ldots, n$, is $\binom{n}{\lfloor (n/2) \rfloor}$.

Solution.
(a) $k \leq (n-1)/2$ implies $2k \leq n-1$ and therefore $k+1 \leq n-k$. Then

$$\frac{1}{k+1} \geq \frac{1}{n-k}.$$

Therefore

$$\binom{n}{k} = \frac{n-k}{\cdot} \frac{n!}{(n-k-1)!k!} \leq \frac{1}{k+1} \cdot \frac{n!}{(n-k-1)!k!} = \binom{n}{(k+1)!}.$$

(b) $k \geq (n-1)/2$ implies

$$n-k-1 \leq n-(n-1)/2-1 \leq (n-1)/2.$$

Using part a we conclude

$$\binom{n}{n-k-1} \leq \binom{n}{n-k},$$

or equivalently

$$\binom{n}{k+1} \leq \binom{n}{k}.$$

(c) is derived by way of (a) and (b). ■

Exercise 157.
For n and m, where $m \leq n$ we have the following.

$$\sum_{k=0}^{m} \binom{n}{k} (-1)^k = (-1)^m \binom{n-1}{m}.$$

Solution.
Use induction on m. ■

Exercise 158.
Find the sum

$$\sum_{i=1}^{n} i\binom{n}{i}.$$

Solution.
Observe that

$$i\binom{n}{i} = i \cdot \frac{n!}{i!(n-i)!} = n \cdot \frac{(n-1)!}{(i-1)!(n-i)!} = n \cdot \binom{n-1}{i-1}.$$

Therefore

$$\sum_{i=1}^{n} i\binom{n}{i}. = \sum_{i=1}^{n} n \cdot \binom{n-1}{i-1} = n \cdot \sum_{i=1}^{n} \binom{n-1}{i-1} = n \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} = n \cdot 2^{n-1},$$

utilizing a previously proven identity. ■

Exercise 159.
From [4]. Let $b \leq b+x < a$ and $0 \leq y < b \leq a$. The following hold.

$$\left(\frac{a-b-x}{a-x}\right)^x \leq \binom{a-x}{b}\binom{a}{b}^{-1} \leq \left(\frac{a-b}{a}\right)^x \leq e^{-(b/a)x}.$$

The following also holds.

$$\left(\frac{b-y}{a-y}\right)^y \leq \binom{a-y}{b-y}\binom{a}{b}^{-1} \leq \left(\frac{b}{a}\right)^y \leq e^{-(1-b/a)y}.$$

Solution.
See [4]. ■

Exercise 160.
(a) For any integer $n > 0$ show the following.

$$(n/e)^n \leq n! \leq en(n/e)^n.$$

(b) For any integer $n > 0$ and $1 \leq k \leq n$ show the following.

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k.$$

Solution.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \\
&= \frac{n^k}{k!} \quad \text{Stirling's approximation} \\
&\leq \frac{n^k}{(k/e)^k} \\
&= \left(\frac{en}{k}\right)^k
\end{aligned}$$

Instead of using Stirling's approximation for the factorial we could have used the following Taylor expansion for e^k.

$$e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} \geq \sum_{i=k}^{i=k} \frac{k^i}{i!} \geq \frac{k^k}{k!}$$

The last one implies $e^k/k^k \geq 1/k!$.
For the lower bound we have the following.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k\cdot(k-1)\cdot\ldots\cdot 1} = \prod_{t=0}^{k-1}\frac{n-t}{k-t} \geq \prod_{t=0}^{k-1}\frac{n}{k} \geq \left(\frac{n}{k}\right)^k
\end{aligned}$$

For $t = 0$, $(n-t)/(k-t) = n/k \geq n/k$. For $t = 1$ through $t = k-1$, we have that $(n-t)/(k-t) \geq n/k$, as long as $k \leq n$. ■

Exercise 161.
Show that the following holds.

$$\binom{2n}{n} = \frac{(2n)!}{n! \cdot n!} \geq 2^n. \tag{4.8}$$

Solution.

$$\begin{aligned}
\frac{(2n)!}{n! \cdot n!} &= \frac{(2n) \cdot (2n-1) \cdot \ldots \cdot (n+1)}{n \cdot (n-1) \cdot \ldots \cdot 1} \\
&= \frac{(2n) \cdot (2n-1) \cdot \ldots \cdot (2n-i) \cdot \ldots \cdot (2n-(n-1))}{n \cdot (n-1) \cdot \ldots \cdot (n-i) \cdot \ldots \cdot (n-(n-1))} \\
&= \frac{(2n)}{n} \cdot \frac{(2n-1)}{n-1} \cdot \ldots \cdot \frac{(2n-i)}{n-i} \cdot \ldots \cdot \frac{2n-(n-1)}{n-(n-1)} \\
&\geq 2 \cdot \frac{n}{n} \cdot 2 \cdot \frac{(n-1/2)}{n-1} \cdot \ldots \cdot 2 \cdot \frac{(n-i/2)}{n-i} \cdot \ldots \cdot 2 \cdot \frac{n-(n-1)/2}{n-(n-1)} \\
&\geq 2^n
\end{aligned}$$

The latter is due to the fact that

$$\frac{n-i/2}{n-i} \geq 1 \Leftrightarrow n - i/2 \geq n - i \Leftrightarrow i/2 \leq i.$$

■

4.6 Fundamental inequalities

Exercise 162.

Show that for $a, b \in \mathbb{R}$ we have the following.

$$ab \leq \frac{1}{2}(a^2 + b^2).$$

AM-GM $n = 2$ case. Moreover, if in addition $a, b \geq 0$, we have the following.

$$\sqrt{ab} \leq \frac{1}{2}(a + b).$$

Solution.

This is equivalent to $(a-b)^2 \geq 0$, which is obviously true. Moreover, the second can be proven by substituting $\sqrt{a}, \sqrt{b}$ for a, b respectively. The latter inequality is known as the geometric arithmetic mean and can be generalized for arbitrary $a_1, \ldots, a_n$ instead of a, b, provided $a_i \geq 0$ and $a_i \in \mathbb{R}$.

■

Exercise 163.
For all $a,b,c,\in\mathbb{R}$ show the following.

$$a^2+b^2+c^2\geq ab+bc+ca.$$

Solution.
In

$$(x-y)^2\geq 0\Rightarrow x^2+y^2\geq 2xy,$$

substitute $x=a,y=b$, $x=b,y=c$ and $x=c,y=a$, and then add the inequalities. ■

Exercise 164.
For all $a,b,c,\in\mathbb{R}$ be non-negative. Show the following.

$$(a+b)(b+c)(c+a)\geq 8abc$$

Solution.
In

$$(x-y)^2\geq 0\Rightarrow x^2+y^2\geq 2xy,$$

substitute $x=\sqrt{a},y=\sqrt{b}$, $x=\sqrt{b},y=\sqrt{c}$ and $x=\sqrt{c},y=\sqrt{a}$, and then multiply the inequalities. ■

Exercise 165.
Let a_i, $i = 1, \ldots, n$ be positive real numbers with $\prod_i a_i = 1$. Show the following.

$$(1+a_1)(1+a_2)\ldots(1+a_n) \geq 2^n.$$

Solution.
In

$$(x-y)^2 \geq 0 \Rightarrow x^2 + y^2 \geq 2xy,$$

set $x = 1$ and $y = a_i$ for all $i = 1, \ldots, n$, and then multiply the inequalityes (all terms positive), noting $\prod_i a_i = 1$. ■

Exercise 166.

Cauchy-Schwartz $n = 2$ case. Show that $(a^2+b^2)(c^2+d^2) \geq (ac+bd)^2$.

Solution.

$$\begin{array}{rcl}
(a^2+b^2)(c^2+d^2) & \geq & (ac+bd)^2 \Leftrightarrow \\
a^2c^2+a^2d^2+b^2c^2+b^2d^2 & \geq & (a^2c^2+b^2d^2+2abcd \Leftrightarrow \\
a^2d^2+b^2c^2+ & \geq & (a^2c^2+b^2d^2+2abcd \Leftrightarrow \\
a^2d^2+b^2c^2-2abcd & \geq & 0 \Leftrightarrow \\
(ad-bc)^2 & \geq & 0 \Leftrightarrow \\
 & T &
\end{array}$$

The last inequality is true. ■

Exercise 167.

Bernoulli inequality. For all $x > -1$ and positive integer n the following is true.

$$(1+x)^n \geq 1+nx$$

Solution.
It is obviously true for $x = 0$ and any n.
It is obviously true for $n = 0$ and any x.
It is also obviously true for $n = 1$.
We omit the proof for other easy cases such ash $0 > x > -1$.
Thus we focus for the $x \geq 0$ case. We show the result by induction on n. For $n = 1$ it is obviously true. If it is true for n, that is

$$(1+x)^n \geq 1+nx$$

then for $n+1$ we have, using the inductive hypothesis, above, and utilizing the fact $nx^2 >= 0$,

$$(1+x)^{n+1} = (1+x)^n(1+x) \geq (1+nx)(1+x) = 1+(n+1)x+nx^2 \geq 1+(n+1)x.$$

■

Exercise 168.

Generalized Bernoulli. Let x_i, $i = 1,\ldots,n$ be real numbers with the same sign and $x_i > -1$. Then the following is true.

$$(1+x_1)(1+x_2)\ldots(1+x_n) \geq 1+x_1+\ldots+x_n.$$

The previous result is for $x_i = x$.

Solution.
Proof by induction. For $n = 1$ obviously $1+x_1 \geq 1+x_1$. Let that for $n = k$ the Bernoulli inequality holds for x_i, $i = 1,\ldots,k$, $x_i > -1$ and all x_i have the same sign and in addition

$$(1+x_1)(1+x_2)\ldots(1+x_k) \geq 1+x_1+\ldots+x_k.$$

Then for $n = k+1$, $x_{k+1} > -1$ and x_{k+1} having the same sign as all the previous x_i we have

$$(x_1+\ldots+x_k)x_{k+1} \geq 0$$

$$(1+x_1)(1+x_2)\ldots(1+x_k)(1+x_{k+1}) \geq (1+x_1+\ldots+x_k)(1+x_{k+1}) \geq 1+x_1+\ldots+x_k+x_{k+1})+(x_1+\ldots+x_k)x_{k+1} \geq 1+x_1+\ldots+x_k+x_{k+1}$$

■

Exercise 169.

(a) **Schur's inequality.** For any $a,b,c,>0$ and positive n show that the following is true.

$$a^n(a-b)(a-c)+b^n(b-c)(b-a)+c^n(c-a)(c-b)\geq 0.$$

Equality is true for $a=b=c$ or two are equal and the third is zero.

(b) Show the following where a,b,c are the sides of a triangle.

$$a^2(b+c-a)+b^2(c+a-b)+c^2(a+b-c)\leq 3abc$$

Solution.

(a) Since inequality is supposed to be true for all a,b,c, then it will be so for a permutation of a,b,c such as $a\leq b\leq c$.

$$a^n(a-b)(a-c)+b^n(b-c)(b-a)+c^n(c-a)(c-b)=(c-b)\left(c^n(c-a)-b^n(b-a)+a^n(a-b)(a-c)\right)\geq 0.$$

The last is so since $c-b\geq 0$, $(a-b)(a-c)\geq 0$, $c-a)\geq b-a$ and $c^n\geq b^n$.

(b) Rewrite the inequality as follows first.

$$a^2b+a^2c+b^2c+b^2a+c^2a+c^2b\leq a^3+b^3+c^3+3abc.$$

Then rearrange the terms, for form Schur's inequality with $n=1$.

$$a(a-b)(a-c)+b(b-c)(b-a)+c(c-a)(c-b)\geq 0.$$

Note about triangles. If a,b,c are the sides (lengths of) a triangle then there exist x,y,z such that $a=x+y$, $b=y+z$, $c=z+x$. Let $s=(a+b+c)/2$. Then $z=s-a,y=s-c,x=s-b$. Moreover the area A of the triangle is $A=(s(s-a)(s-b)(s-c))^{1/2}$. ■

Exercise 170.
(a) For any non-negative $a,b,c \in \mathbb{R}$ show the following

$$\frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b} \geq \frac{3}{2}.$$

Solution.
(a) Form

$$S = \frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b},$$

and then

$$T = \frac{b}{b+c}+\frac{c}{c+a}+\frac{a}{a+b},$$

$$U = \frac{c}{b+c}+\frac{a}{c+a}+\frac{b}{a+b},$$

By inspection $T+U=3$. Furthermore, using the AG-mean

$$S+T = \frac{a+b}{b+c}+\frac{b+c}{c+a}+\frac{c+a}{a+b} \geq 3\cdot 1,$$

and

$$S+U = \frac{a+c}{b+c}+\frac{b+a}{c+a}+\frac{c+b}{a+b} \geq 3\cdot 1.$$

Therefore

$$S+T+S+U \geq 6 \Leftrightarrow 2S \geq 3 \Leftrightarrow S \geq \frac{3}{2}.$$

■

Exercise 171.

Arithmetic-Geometric (AG) mean. Let $a_1,\ldots,a_n$ be non-negative real numbers. Then show the following.

$$\sqrt[n]{a_1a_2\ldots a_n} \leq \frac{a_1+a_2+\ldots+a_n}{n}.$$

Solution.
Proof by induction. The $n=2$ case follows immediately from $(\sqrt{a_1}-\sqrt{a_2})^2 \geq 0$.
We then use induction on k for $n/2=2^{k-1}$ to $n=2^k$ with $k>1$. So let $b_1,\ldots,b_{n/2}$ are such that,

$$\sqrt[n/2]{b_1b_2\ldots b_{n/2}} \leq \frac{b_1+b_2+\ldots+b_{n/2}}{n/2}.$$

The sequence $b_1,\ldots,b_{n/2}$ will be replaced by the induction hypothesis by $a_1,\ldots,a_{n/2}$ or by $a_{n/2+1},\ldots,a_n$ (step 2 to step 3). We apply the $n=2$ case for c_1 and c_2 defined as follows.

$$c_1=\frac{a_1+\ldots+a_{n/2}}{n/2}, \qquad c_2=\frac{a_{n/2+1}+\ldots+a_n}{n/2}.$$

$$\sqrt{c_1c_2} \leq \frac{c_1+c_2}{2}.$$

By the induction hypothesis (step 2 to step 3 below)

$$\begin{aligned}
\frac{c_1+c_2}{2} &= \frac{1}{2}\cdot\frac{a_1+a_2+\ldots+a_n}{n/2} \geq \sqrt{c_1c_2} \Leftrightarrow \\
\frac{a_1+a_2+\ldots+a_n}{n} &\geq \sqrt{\frac{a_1+\ldots+a_{n/2}}{n/2}}\sqrt{\frac{a_{n/2+1}+\ldots+a_n}{n/2}} \Leftrightarrow \\
\frac{a_1+a_2+\ldots+a_n}{n} &\geq (a_1\ldots a_{n/2})^{\frac{1}{n}}(a_{n/2+1}\ldots a_n)^{\frac{1}{n}} \Leftrightarrow \\
\frac{a_1+a_2+\ldots+a_n}{n} &\geq (a_1a_2\ldots a_n)^{\frac{1}{n}}.
\end{aligned}$$

This proves the validity of the inequality for values of n that are powers of two. What if n is such that $2^{k-1}<n<2^k$? We define $S=a_1+a_2+\ldots+a_n$. We generate a sequence whose length is NOT a power of two now given that n is such that $2^{k-1}<n<2^k$. and extend that sequence to one that is a power of two to apply the inductive hypothesis!

$$b_1=a_1,\ldots,b_n=a_n, \quad b_{n+1}=\ldots=b_{2^k}=\frac{a_1+a_2+\ldots+a_n}{n}=\frac{S}{n}.$$

Using the prior results for 2^k we then have for $b_1,\ldots,b_{2^k}$ the following.

$$\begin{aligned}
\frac{b_1+\ldots+b_{2^k}}{2^k}=\frac{S+S(2^k-n)/n}{2^k} &\geq \left(a_1\ldots a_n(S/n)^{2^k-n}\right)^{1/2^k} \Leftrightarrow \\
\frac{S}{n} &\geq (a_1\ldots a_n)^{1/2^k}\,\frac{S^{1-n/2^k}}{n^{1-n/2^k}} \Leftrightarrow \\
\frac{S^{n/2^k}}{n} &\geq (a_1\ldots a_n)^{1/2^k}\,\frac{n^{n/2^k}}{n} \Leftrightarrow \\
S^{n/2^k} &\geq (a_1\ldots a_n)^{1/2^k}\,n^{n/2^k} \Leftrightarrow \\
S^n &\geq (a_1\ldots a_n)\,n^n \Leftrightarrow \\
\frac{S}{n} &\geq (a_1\ldots a_n)^{\frac{1}{n}} \Leftrightarrow \\
\frac{a_1+a_2+\ldots+a_n}{n} &\geq (a_1\ldots a_n)^{\frac{1}{n}}.
\end{aligned}$$

■

Exercise 172.

Arithmetic-Geometric (AG) mean. Let $a_1, \ldots, a_n$ be non-negative real numbers. Then show the following.

$$\sqrt[n]{a_1 a_2 \ldots a_n} \leq \frac{a_1 + a_2 + \ldots + a_n}{n}.$$

Solution.
(Proof of G. Polya.)
Consider function $f(x) = e^{x-1} - x$. Then $f'(x) = e^{x-1} - 1$. Then $f''(x) = e^{x-1}$. We observe $f(1) = 0 = f'(1)$ and $f''(x) > 0$ for all x. Function f is convex, and has a minimum at $x = 1$. Therefore $f(x) \geq f(1)$ and then $e^{x-1} \geq x$.
Case 1: $a_i = 0, \forall i$**.** Then the AG mean is true (equality holds).
Case 2: $\exists k : a_k \neq 0$**.** Then since all a_i are non-negative integers, and $a_k \neq 0$, the $\sum_i a_i/n = a > 0$. In

$$e^{x-1} \geq x,$$

substitue $a_1/a, a_2/a, \ldots, a_n/a$ for x, and multiply all the inequalities. We derive the following.

$$\begin{aligned}
e^{a_1/a-1} \cdot e^{a_2/a-1} \cdot \ldots \cdot e^{a_n/a-1} &\geq \frac{a_1}{a} \cdot \frac{a_2}{a} \cdot \ldots \cdot \frac{a_n}{a} \Leftrightarrow \\
e^{\sum_i (a_i/a)-n} &\geq \frac{a_1 a_2 \ldots a_n)}{a^n} \\
e^{na/a-n} &\geq \frac{a_1 a_2 \ldots a_n}{a^n} \\
1 &\geq \frac{(a_1 a_2 \ldots a_n)}{a^n} \\
\frac{a_1 + a_2 + \ldots + a_n}{n} &\geq \sqrt[n]{a_1 a_2 \ldots a_n}.
\end{aligned}$$

Note that Polya used in his proof $f(x) = e^x - x - 1$, then generating a change of variable from x to $x - 1$. ■

Exercise 173.

The other Bernoulli inequality. For all $x > -1$ and positive rational number a such that $0 < a < 1$ the following is true.

$$(1+x)^a \leq 1+ax$$

Solution.

If a is rational, there exist n,m such that $a = n/m$ and $1 \leq n < m$. Since $x > -1$ then $x+1 \geq 0$. We apply the AG-mean.

$$\begin{aligned}
(1+x)^a &= (1+x)^{\frac{n}{m}} \\
&= \sqrt[m]{(1+x)^n} \\
&= \sqrt[m]{\underbrace{(1+x)(1+x)\ldots(1+x)}_{n \text{ terms}}\underbrace{1\cdot 1\cdot \ldots \cdot 1}_{m-n \text{ term}}} \\
&\leq \frac{(1+x)+\ldots+(1+x)+(m-n)}{m} \\
&\leq \frac{(1+x)n+(m-n)}{m} \\
&\leq 1+x\frac{n}{m} \\
&\leq 1+xa
\end{aligned}$$

■

Exercise 174.

Cauchy-Schwartz general case. For any real $a_1, a_2, \ldots, a_n \mathbb{R}$, and any real $b_1, b_2, \ldots, b_n \mathbb{R}$, show the following.

$$(a_1^2 + a_2^2 + \ldots + a_n^2)(b_1^2 + b_2^2 + \ldots + b_n^2) \geq (a_1 b_1 + a_2 b_2 + \ldots + a_n b_n)^2.$$

Equality applies for $a_i/b_i = a_j/b_j$ for all $i, \neq j$.
Note that for $a_i = \sqrt{A_i}$, $b_i = \sqrt{B_i}$, $c_i = \sqrt{C_i}$, the inequality takes the following form.

$$(A_1 + A_2 + \ldots + A_n)^{1/2}(B_1 + B_2 + \ldots + B_n)^{1/2} \geq ((A_1 B_1)^{1/2} + (A_2 B_2)^{1/2} + \ldots + \sqrt{A_n B_n}^{1/2}.$$

Solution.
From

$$(x-y)^2 \geq 0 \Leftrightarrow x^2 + y^2 \geq 2xy,$$

substitute

$$x = \frac{a_i}{\sqrt{\sum_i a_i^2}}, \quad y = \frac{b_i}{\sqrt{\sum_i b_i^2}}, \quad \forall i = 1, \ldots, n,$$

and them multiply them together. One derives the following.

$$\sum_i \frac{a_i^2}{\sum_i a_i^2} + \sum_i \frac{b_i^2}{\sum_i b_i^2} \geq \frac{\sum_i (2a_i b_i)}{\sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2}},$$

which is equivalent to

$$2 \geq 2 \cdot \frac{\sum_i (a_i b_i)}{\sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2}},$$

and after diving by two and squaring, the result then follows.
An alternative proof. Consider

$$f(x) = \sum_i (a_i x - b_i)^2.$$

It is clear that $f(x) \geq 0$ with $f(x) = 0$ if $a_i x = b_i$ for all $i = 1, \ldots, n$ for some $x \in \mathbb{R}$. $f(x)$ is a quadratic polynomial in x of the form $f(x) = ax^2 + bx + c$. We note that $a = \sum_i a_i^2$, $c = \sum_i b_i^2$. Since $f(x) \geq 0$ the discriminant $\Delta = b^2 - 4ac \leq 0$, and the result follows. ■

Exercise 175.
(a) Prove the following for $a,b,c \in \mathbb{R}$.

$$3(a^2+b^2+c^2) \geq (a+b+c)^2$$

(b) Prove the following for $a,b,c \in \mathbb{R}$ and positive.

$$\frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b} \geq 1.5.$$

Solution.
(a) Set $a_i = 1$, $b_1 = a, b_2 = b, b_3 = c$ and use the Cauchy-Schwartz general case for $n = 3$.
Add 3 to both sides.

$$\frac{a}{b+c}+1+\frac{b}{c+a}+1+\frac{c}{a+b}+1 \geq 4.5 \Leftrightarrow \frac{a}{b+c}+\frac{b+c}{b+c}+\frac{b}{c+a}+\frac{a+c}{a+c}+\frac{c}{a+b}+\frac{a+b}{a+b} \geq 4.5$$

One then needs to show the following.

$$(a+b+c)\left(\frac{1}{b+c}+\frac{1}{c+a}+\frac{1}{a+b}\right) \geq 4.5 \Leftrightarrow (2a+2b+2c)\left(\frac{1}{b+c}+\frac{1}{c+a}+\frac{1}{a+b}\right) \geq (1+1+1)^2.$$

Set $x = \sqrt{a+b}$, set $y = \sqrt{b+c}$, set $z = \sqrt{c+a}$, and then use the Cauchy-Schwartz general case for $n = 3$, for x,y,z and $1/x, 1/y, 1/z$. ■

Exercise 176.

Holder's inequality for $n=3$. Let $a_1,a_2,a_3,b_1,b_2,b_3,c_1,c_2,c_3$ be positive real numbers with $a_i,b_i,c_i \in \mathbb{R}, i = 1,2,3..$ The following applies.

$$(a_1^3+a_2^3+a_3^3)(b_1^3+b_2^3+b_3^3)(c_1^3+c_2^3+c_3^3) \geq (a_1b_1c_1+a_2b_2c_2+a_3b_3c_3)^3.$$

An alternative form is as follows. Let $A_1,A_2,A_3,B_1,B_2,B_3,C_1,C_2,C_3$ be positive real numbers with $A_i,B_i,C_i \in \mathbb{R}, i = 1,2,3..$ Show the following

$$(A_1+A_2+A_3)(B_1+B_2+B_3)(C_1+C_2+C_3) \geq \left((A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3}\right)^3.$$

Set $a_i^3 = A_i$, $b_i^3 = B_i$, and $c_i^3 = C_i$, and the result follows from the previous one.
Holder's inequality can be generalized to n from $n=3$, and also to n and m with m different sequences of length n.

Solution.
Proof follows from the AG-mean. Let us define the following expressions D_1,D_2,D_3 as follows.

$$D_1 = \frac{(A_1B_1C_1)^{1/3}}{(A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3}} = \frac{(A_1B_1C_1)^{1/3}}{D} \quad \Rightarrow \quad D = \frac{(A_1B_1C_1)^{1/3}}{D_1},$$

$$D_2 = \frac{(A_2B_2C_2)^{1/3}}{(A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3}} = \frac{(A_2B_2C_2)^{1/3}}{D} \quad \Rightarrow \quad D = \frac{(A_2B_2C_2)^{1/3}}{D_2},$$

$$D_3 = \frac{(A_3B_3C_3)^{1/3}}{(A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3}} = \frac{(A_3B_3C_3)^{1/3}}{D} \quad \Rightarrow \quad D = \frac{(A_3B_3C_3)^{1/3}}{D_3},$$

where

$$D = (A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3}.$$

It is obvious that $D_1+D_2+D_3 = 1$. Consider $A_1+A_2+A_3$ (and then repeat for B_i,C_i). We will apply AG-mean result to D_1 instances of A_1/D_1, D_2 instances of A_2/D_2, and D_3 instances of A_3/D_3 noting also that $D_1+D_2+D_3 = 1$.

$$\begin{aligned} A_1+A_2+A_3 &= \frac{A_1}{D_1}D_1+\frac{A_2}{D_2}D_2+\frac{A_3}{D_3}D_3 \\ &\geq \sqrt[D_1+D_2+D_3]{\left(\frac{A_1}{D_1}\right)^{D_1}\left(\frac{A_2}{D_2}\right)^{D_2}\left(\frac{A_3}{D_3}\right)^{D_3}} \\ &= \left(\frac{A_1}{D_1}\right)^{D_1}\left(\frac{A_2}{D_2}\right)^{D_2}\left(\frac{A_3}{D_3}\right)^{D_3} \end{aligned}$$

Similarly for B_i,C_i, the following can be obtained.

$$(B_1+B_2+B_3) \geq \left(\frac{B_1}{D_1}\right)^{D_1}\left(\frac{B_2}{D_2}\right)^{D_2}\left(\frac{B_3}{D_3}\right)^{D_3}, \quad (C_1+C_2+C_3) \geq \left(\frac{C_1}{D_1}\right)^{D_1}\left(\frac{C_2}{D_2}\right)^{D_2}\left(\frac{C_3}{D_3}\right)^{D_3}.$$

Multiply the three previous results and then raise them to the $1/3$ power.

$$(A_1+A_2+A_3)^{1/3}(B_1+B_2+B_3)^{1/3}(C_1+C_2+C_3)^{1/3} \quad \geq \quad \left(\frac{A_1B_1C_1}{D_1^3}\right)^{D_1/3}\left(\frac{A_2B_2C_2}{D_2^3}\right)^{D_2/3}\left(\frac{A_3B_3C_3}{D_3^3}\right)^{D_3/3}$$

Isolate one of the three terms of the right hand side and use Eq.(4.9), Eq.(4.9), and Eq.(4.9).

$$\left(\frac{A_1B_1C_1}{D_1^3}\right)^{D_1/3}=\left(\left(\frac{A_1B_1C_1}{D_1^3}\right)^{1/3}\right)^{D_1}=\left(\frac{(A_1B_1C_1)^{1/3}}{D_1}\right)^{D_1}=D^{D_1},$$

and similarly,

$$\left(\frac{A_2B_2C_2}{D_2^3}\right)^{D_2/3}=D^{D_2},\left(\frac{A_3B_3C_3}{D_3^3}\right)^{D_3/3}=D^{D_3}.$$

Eq.(4.9) by way of the three last results gives.

$$(A_1+A_2+A_3)^{1/3}(B_1+B_2+B_3)^{1/3}(C_1+C_2+C_3)^{1/3}\geq D^{D_1+D_2+D_3}=D.$$

Therefore we just proved the following

$$(A_1+A_2+A_3)^{1/3}(B_1+B_2+B_3)^{1/3}(C_1+C_2+C_3)^{1/3}\geq (A_1B_1C_1)^{1/3}+(A_2B_2C_2)^{1/3}+(A_3B_3C_3)^{1/3},$$

as needed, since if we raise both sides to the third power the result is obtained. ■

Exercise 177.

Jensen's inequality. (a) Let f be a convex function in the closed interval $[a,b]$. Let $n \in \mathbb{N}$ Then the inequality is true for all $x_i \in [a,b]$).

$$f\left(\frac{x_1+x_2+\ldots+x_n}{n}\right) \leq \frac{f(x_1)+f(x_2)+\ldots+f(x_n)}{n}.$$

For a concave function the inequality flips.

Since f is convex $f''(x) \geq 0$ for all $x \in [a,b]$.

(b) For $a,b,c > 0$ show the following

$$a^a b^b c^c \geq \left(\frac{a+b+c}{3}\right)^{a+b+c}.$$

Solution.

(a) Proof by induction. We shall prove a more general result in particular the following

$$f\left(\sum_{i=1}^{n} a_i x_i\right) \leq \sum_{i=1}^{n} a_i f(x_i),$$

where $\sum_{i=1}^{n} a_i = 1$, $n \in \mathbb{N}$, and $a_i \in (0,1)$, $a_i \in \mathbb{R}$, with $x_i \in [a,b]$. Note that in the form depicted in (a) above $a_i = 1/n$. Base case $n = 1$. $f(x_1) = f(x_1)$ and for $a_1 = 1$ we have $f(a_1 x_1) = a_1 f(x_1)$ and the case is resolved. Induction hypothesis. Suppose that for $n = k$ and for $a_1 + a_2 + \ldots + a_k = 1$, $a_i \in (0,1)$, $a_i \in \mathbb{R}$, $x_i \in [a,b]$ we have

$$f(a_1 x_1 + a_2 x_2 + \ldots + a_k x_k) \leq a_1 f(x_1) + a_2 f(x_2) + \ldots + a_k f(x_k).$$

Consider $n = k+1$ and let $a_1 + a_2 + \ldots + a_k + a_{k+1} = 1$, $a_i \in (0,1)$, $a_i \in \mathbb{R}$ then let $x_1, \ldots, x_{k+1} \in [a,b]$, and the following expression.

$$\begin{aligned} a_1 x_1 + \ldots + a_k x_k + a_{k+1} x_{k+1} &= (a_1 x_1 + a_2 x_2 + \ldots + a_k x_k) + a_{k+1} x_{k+1} \\ &= (1 - a_{k+1}) \left(\frac{a_1}{1-a_{k+1}} x_1 + \frac{a_2}{1-a_{k+1}} x_2 + \ldots + \frac{a_k}{1-a_{k+1}} x_k\right) + a_{k+1} x_{k+1} \quad (4.9) \end{aligned}$$

Note that

$$\frac{a_1}{1-a_{k+1}} + \frac{a_2}{1-a_{k+1}} + \ldots + \frac{a_k}{1-a_{k+1}} = \frac{a_1 + a_2 + \ldots a_k}{1-a_{k+1}} = \frac{1-a_{k+1}}{1-a_{k+1}} = 1. \quad (4.10)$$

Let

$$z = \frac{a_1}{1-a_{k+1}} x_1 + \frac{a_2}{1-a_{k+1}} x_2 + \ldots + \frac{a_k}{1-a_{k+1}} x_k$$

Sice $x_i \in [a,b]$ we have $a \leq x_i \leq b$. It is obvious then that $a \leq z \leq b$. For example (the case for b is similar).

$$z = \frac{a_1}{1-a_{k+1}} x_1 + \frac{a_2}{1-a_{k+1}} x_2 + \ldots + \frac{a_k}{1-a_{k+1}} x_k \geq \frac{a_1}{1-a_{k+1}} a + \frac{a_2}{1-a_{k+1}} a + \ldots + \frac{a_k}{1-a_{k+1}} a \geq \frac{a_1 + a_2 + \ldots a_k}{1-a_{k+1}} \cdot a = a.$$

By the induction hypothesize for $n = k$ we have for z and $f(z)$ below given Eq.(4.10).

$$\begin{aligned} f(z) &= f\left(\frac{a_1}{1-a_{k+1}} x_1 + \frac{a_2}{1-a_{k+1}} x_2 + \ldots + \frac{a_k}{1-a_{k+1}} x_k\right) \\ &\leq \frac{a_1}{1-a_{k+1}} f(x_1) + \frac{a_2}{1-a_{k+1}} f(x_2) + \ldots + \frac{a_k}{1-a_{k+1}} f(x_k). \quad (4.11) \end{aligned}$$

By way of Eq.(4.9) we have the following.

$$\begin{aligned} f(a_1x_1+a_2x_2+\ldots+a_kx_k+a_{k+1}x_{k+1}) &= f((a_1x_1+a_2x_2+\ldots+a_kx_k)+a_{k+1}x_{k+1}) \\ &= f((1-a_{k+1})\left(\frac{a_1}{1-a_{k+1}}x_1+\frac{a_2}{1-a_{k+1}}x_2+\ldots+\frac{a_k}{1-a_{k+1}}x_k\right)+a_{k+1}x_{k+1}) \\ &= f((1-a_{k+1})z+a_{k+1}x_{k+1}) \\ &\leq (1-a_{k+1})f(z)+a_{k+1}f(x_{k+1}). \end{aligned} \tag{4.12}$$

Equation Eq.(4.12) by way of Eq.(4.11) yields the following.

$$\begin{aligned} f(a_1x_1+a_2x_2+\ldots+a_kx_k+a_{k+1}x_{k+1}) &\leq (1-a_{k+1})f(z)+a_{k+1}f(x_{k+1}) \\ &\leq (1-a_{k+1})\left(\frac{a_1}{1-a_{k+1}}f(x_1)+\frac{a_2}{1-a_{k+1}}f(x_2)+\ldots+\frac{a_k}{1-a_{k+1}}f(x_k)\right)+a_{k+1}f(x_{k+1}) \quad (4.13) \\ &\leq a_1f(x_1)+\ldots+a_{k+1}f(x_{k+1}), \end{aligned} \tag{4.14}$$

as needed to complete the inductive step.
(b) Consider $f(x)=x\ln x$. Then $f'(x)=\ln x+1$. Therefore, $f''(x)=1/x$. Thus for $x\in(0,\infty)$, we have that $f''(x)>0$ and f is convex. Thefore for $a,b,c\in(0,\infty)$ we have

$$\begin{aligned} 3f((a+b+c)/3) &\leq f(a)+f(b)+f(c) \\ &\leq a\ln a+b\ln b+c\ln c \\ &\leq \ln(a^ab^bc^c)\Leftrightarrow \\ 3\frac{a+b+c}{3}\cdot\ln\frac{a+b+c}{3} &\leq \ln(a^ab^bc^c)\Leftrightarrow \\ \left(\frac{a+b+c}{3}\right)^{a+b+c} &\leq a^ab^bc^c. \end{aligned}$$

■

Exercise 178.

Arithmetic-Geometric (AG) mean. Let $x_1, \ldots, x_n$ be non-negative real numbers. Then show the following using Jensen's inequality.

$$\sqrt[n]{x_1 x_2 \ldots x_n} \leq \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Solution.

Consider $f(x) = -\ln(x)$ defined on the open interval $(0, \infty)$. Then $f'(x) = -\frac{1}{x}$, and $f''(x) = \frac{1}{x^2}$. Moreover $f''(x) > 0$. Therefore f is convex (in fact strictly convex) on $(0, \infty)$. Use Jensen's inequality with $a_i = 1/n$ for each $i = 1, \ldots, n$. We then have

$$\begin{aligned}
-\ln\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) &\leq -\left(\frac{\ln x_1 + \ln x_2 + \ldots + \ln x_n}{n}\right) \Leftrightarrow \\
\left(\frac{\ln x_1 + \ln x_2 + \ldots + \ln x_n}{n}\right) &\leq \ln\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) \Leftrightarrow \\
\ln\left(\sqrt[n]{x_1 x_2 \ldots x_n}\right) &\leq \ln\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right),
\end{aligned}$$

and the result follows after raising e to the exponents of the left and the right-hand side above. ■

Exercise 179.

Young's inequality. Let $X,Y,p,q \in \mathbb{R}$ with $X,Y>0$ and $p,q>1$, so that $1/p+1/q=1$. The the following holds

$$XY \leq \frac{X^p}{p}+\frac{Y^q}{q}.$$

Equality is for $X^p = Y^q$.

Solution.
Consider convex function $f(x)=e^x$ for any $x \in \mathbb{R}$. Naturally f is convex on $(0,\infty)$. Since $1/p+1/q=1$ uses Jensen's inequality for $n=2$ as follows. Use $x=p\ln X$ and $y=q\ln Y$.

$$\begin{aligned} f(\frac{x}{p}+\frac{y}{q}) &\leq \frac{1}{p}f(x)+\frac{1}{q}f(y) \Leftrightarrow \\ e^{\frac{x}{p}+\frac{y}{q}} &\leq \frac{1}{p}e^x+\frac{1}{q}e^y \Leftrightarrow \\ e^{\ln X+\ln Y} &\leq \frac{1}{p}e^{p\ln X}+\frac{1}{q}e^{q\ln Y} \Leftrightarrow \\ XY &\leq \frac{X^p}{p}+\frac{Y^q}{q}. \end{aligned}$$

■

Exercise 180.

Holder's general inequality. Let $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$ be positive real numbers with $a_i, b_i \in \mathbb{R}$, for all $i = 1, 2, \ldots, n$. Let $p, q \in \mathbb{R}$ suchthat $p, q > 1$ and $1/p + 1/q = 1$. The following applies.

$$\left(\sum_{i=1}^{n} a_i b_i\right) \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}.$$

Solution.
Use Young's inequality with repeated substitutions $X = X_i$ and $Y = Y_i$ for $i = 1, 2, \ldots, n$.

$$X_i = \frac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}}, \quad Y_i = \frac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}}.$$

Then

$$X_i Y_i = \frac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} \cdot \frac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}} \leq \frac{a_i^p}{p \cdot \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} + \frac{b_i^q}{q \cdot \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}}$$

Then

$$\sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} \left(\frac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} \cdot \frac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}} \right) \leq \frac{\sum_i a_i^p}{p \cdot \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} + \frac{\sum_i b_i^q}{q \cdot \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}} = \frac{1}{p} + \frac{1}{q} = 1,$$

i.e.

$$\sum_{i=1}^{n} \left(\frac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} \cdot \frac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}} \right) \leq 1 \Leftrightarrow \sum_{1=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}.$$

■

Exercise 181.

Minkowski's inequality. Let $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$ be positive real numbers with $a_i, b_i \in \mathbb{R}$, for all $i = 1, 2, \ldots, n$. Let $p \in \mathbb{R}$ such that $p > 1$. The following applies.

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}b_i^q\right)^{\frac{1}{p}}.$$

Equality is for $a_1/b_1 = a_2/b_2 = \ldots = a_n/b_n$.

Solution.
Given $p > 1$ find a $q > 1$ such that $1/p + 1/q = 1$. Solving for q we have, $q = p/(p-1)$. Apply then Holder's inequality.

$$\begin{aligned}
\sum_{i=1}^{n}(a_i+b_i)^p &= \sum_{i=1}^{n}(a_i+b_i)^{p-1}(a_i+b_i) \\
&= \sum_{i=1}^{n}a_i(a_i+b_i)^{p-1} + \sum_{i=1}^{n}b_i(a_i+b_i)^{p-1} \\
&\leq \left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n}(a_i+b_i)^{(p-1)q}\right)^{\frac{1}{q}} + \left(\sum_{i=1}^{n}b_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n}(a_i+b_i)^{(p-1)q}\right)^{\frac{1}{q}} \\
&\leq \left(\sum_{i=1}^{n}(a_i+b_i)^{(p-1)q}\right)^{\frac{1}{q}} \left(\left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}b_i^p\right)^{\frac{1}{p}}\right) \\
&\leq \left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{p-1}{p}} \left(\left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}b_i^p\right)^{\frac{1}{p}}\right) \Leftrightarrow \\
\sum_{i=1}^{n}(a_i+b_i)^p \left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{-\frac{p-1}{p}} &\leq \left(\left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}b_i^p\right)^{\frac{1}{p}}\right) \Leftrightarrow \\
\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{p}} &\leq \left(\left(\sum_{i=1}^{n}a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n}b_i^p\right)^{\frac{1}{p}}\right).
\end{aligned}$$

■

Exercise 182.

(a) **Permuting monotonic sequences.** Let $(x_i)_{i=1}^n$ and let $(y_i)_{i=1}^n$ be sequence of positive real values both increasing or decreasing in the same direction. For any permutation $(z_i)_{i=1}^n$ of y_i show the following.

$$\sum_{i=1}^{n} x_i y_i \geq \sum_{i=1}^{n} x_i z_i \geq \sum_{i=1}^{n} x_i y_{n-i+1}.$$

(b) Next show for $1 \leq a \leq b \leq c$ the following is true.

$$a^a b^b c^c \geq a^b b^c c^a.$$

(A circular shift of the exponents.) It is also then true if $a \geq b \geq c \geq 1$.

Solution.
(a) This result can be proven useful in the proof of correctness of greedy algorithms, e.g. Huffman's compression method. Let

$$A = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

and let B be the sequence where j,k indexed get swapped.

$$B = x_1 y_1 + x_2 y_2 + \ldots + x_j y_k + \ldots + x_k y_j + \ldots + x_n y_n.$$

Consider then $A - B$

$$A - B = x_j y_j + x_k y_k - x_j y_k - x_k y_j = x_j(y_j - y_k) - x_j(y_j - y_k) = (x_j - x_k)(y_j - y_k) \geq 0$$

Thus if the orders of monotonicity of x and y are the same the difference $A - B$ is positive, if not it is negative. Moreover, any time we have a swap, the sum gets smaller. This implies the maximum is the original A whether both sequences are increasing (monotonically non-decreasing) or decreasing (monotonically non-increasing).
(b) Take the logarithms of both sides and use (a) to prove the following instead.

$$\lg a^a b^b c^c = a \lg a + b \lg b + c \lg c \geq \lg a^b b^c c^a = a \lg b + b \lg c + c \lg a$$

The sequence a,b,c has the same monotonicity as $\lg a, \lg b, \lg c$. ■

Exercise 183.

AGH (Arithmetic, Geometric and Harmonic mean inequality) for $n = 2$. Show that for $a, b > 0$ we have

$$(a+b)/2 \geq \sqrt{ab} \geq 2/(1/a+1/b).$$

Solution.
First part first (left-most inequality).
In order to show $(a+b)/2 \geq \sqrt{ab}$ we square both sides. Since all quantities are non-negative, it suffices to show $(a+b)^2/4 \geq ab$ i.e. $(a+b)^2 \geq 4ab$, i.e. $(a+b)^2 - 4ab \geq 0$ that is $(a-b)^2 \geq 0$. Since a, b are non-negative, the latter is true.
Second part next (right-most inequality).
In order to show $\sqrt{ab} \geq 2/(1/a+1/b)$, we need to show the following sequence of equivalent inequalities. The last one is obviously true. Note that the third inequality had both its sides multiplied with $ab > 0$, and the first squaring has both sides greather than zero as well.

$$\begin{aligned}
(\sqrt{ab})^2 &\geq (2/(1/a+1/b))^2 \Leftrightarrow \\
ab &\geq 4/(1/a+1/b)^2 \Leftrightarrow \\
ab(1/a+1/b)^2 &\geq 4 \Leftrightarrow \\
a^2b^2(1/a+1/b)^2 &\geq 4ab \Leftrightarrow \\
a^2+b^2+2ab &\geq 4ab \Leftrightarrow \\
a^2+b^2-2ab &\geq 0 \Leftrightarrow \\
(a-b)^2 &\geq 0 \Leftrightarrow \\
T &
\end{aligned}$$

The last inequality above is true (square of a real number cannot be negative).
First and second inequalities proven. The generalization of this for n terms instead of two is known as the Arithmetic-Geometric-Harmonic Mean inequality. ■

Exercise 184.

Chebyshev's inequality. Let $(x_i)_{i=1}^n$ and let $(y_i)_{i=1}^n$ be sequences of positive real values of the same monotonicity. Show the following

$$\frac{x_1y_1+x_2y_2+\ldots+x_ny_n}{n} \geq \frac{x_1+x_2+\ldots+x_n}{n}\frac{y_1+y_2+\ldots+y_n}{n}$$

Solution.

Use the Permuting monotonic sequences problem after rewriting the left-hand side n different ways.

$$\begin{aligned}
x_1y_1+x_2y_2+\ldots+x_ny_n &= x_1y_1+x_2y_2+\ldots+x_ny_n\\
x_1y_1+x_2y_2+\ldots+x_ny_n &\geq x_1y_2+x_2y_3+\ldots+x_ny_1\\
x_1y_1+x_2y_2+\ldots+x_ny_n &\geq x_1y_3+x_2y_4+\ldots+x_ny_2\\
&\ldots\\
x_1y_1+x_2y_2+\ldots+x_ny_n &\geq x_1y_n+x_2y_1+\ldots+x_ny_{n-1}.
\end{aligned}$$

Add them up and the result follows after a division of both sides with n^2.

$$n\left(x_1y_1+x_2y_2+\ldots+x_ny_n\right) \geq \left(x_1+x_2+\ldots+x_n\right)\left(y_1+y_2+\ldots+y_n\right).$$

■

Exercise 185.
For any $a,b \in R$ and $a \neq b$, we have

$$2ab/(a+b) \leq \sqrt{ab} \leq (a-b)/(\ln(a)-\ln(b)) \leq (a+b)/2.$$

Solution.
The extremes are the Arithmetic-Geometric mean inequality proved elsewhere. We show the following part

$$\sqrt{ab} \leq (a-b)/(\ln(a)-\ln(b)) \leq (a+b)/2.$$

The proof is for

$$\sqrt{ab} \leq (a-b)/(\ln(a)-\ln(b)).$$

The other part is proven similarly. Let us assume, without loss of generality, that $a > b$. Let $z = a/b > 1$ then. The inequality to be proven reduces to the following one.

$$\begin{aligned} \sqrt{ab} &\leq b(a/b-1)/(\ln(a/b)) \\ \sqrt{ab/b^2} &\leq (a/b-1)/(\ln(a/b)) \\ \sqrt{z} &\leq (z-1)/\ln(z). \end{aligned}$$

Given that $z > 1$, let $z = t^2$. Then the previous inequality (to be proven) can be rewritten as follows.

$$\begin{aligned} t &\leq (t^2-1)/2\ln(t) \\ 2\ln(t) &\leq (t^2-1)/t \\ 2\ln(t) &\leq t-1/t. \end{aligned}$$

If $g(t) = 2\ln(t)$ and $h(t) = t-1/t$, we observe $g(1) = h(1) = 0$ and $g' < h'$ since $2/t < 1+1/t^2$ which is equivalent to $(1-1/t)^2 > 0$, a true assertion for $z = a/b > 1$ and $z = t^2$. Moreover one considering $f(t) = t-1/t-2\ln(t)$, and considering $f' = 1+1/t^2-2/t = (1-1/t)^2 > 0$, given that $f(1) = 0$ we conclude $f(t)$ is increasing for $t > 1$ and $f(t) > 0$ proving the $2\ln(t) \leq t-1/t$. ■

4.7 Natural/Mercator logarithm arithmetic

Exercise 186.
Show that $(1-1/N)^n \geq (1-n/N)$ for any integer $n \geq 1$ and $N \geq 1$.

Solution.
Method 1. Proof by induction on n. For $n=1$ obviously $(1-1/N)^1 \geq (1-1/N)$.
For $n=k$ let $(1-1/N)^k \geq 1-k/N$. Then for $n=k+1$ we have

$$\begin{aligned}(1-1/N)^{k+1} &= (1-1/N)^k \cdot (1-1/N) \\ &\geq (1-k/N)\cdot(1-1/N) \\ &= 1-(k+1)/N+k/N^2 \\ &\geq 1-(k+1)/N.\end{aligned}$$

Method 2. An easier proof is by an application of the Bernoulli inequality, by using $x=-1/N$ and $n=n$. (Note that $-1/N > -1$ since $N>1$, and $n \geq 1$.) ■

Exercise 187.
Show that $(1+1/n)^n < 3$ for all positive integer $n > 2$.

Solution.
First we note that for $k > 1$, we have $k! = k \cdot (k-1) \cdot \ldots 3 \cdot 2 \cdot 1 \geq 2 \cdot 2 \cdot \ldots 2 \cdot 2 \cdot 1 = 2^{k-1}$. By the binomial theorem

$$\begin{aligned}
(1+\frac{1}{n})^n &= \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} \\
&= 1+1+\sum_{k=2}^{n} \binom{n}{k} \frac{1}{n^k} \\
&= 1+1+\sum_{k=2}^{n} (1-\frac{1}{n}) \ldots (1-\frac{k-1}{n}) \frac{1}{k!} \\
&\leq 2+\sum_{k=2}^{n} \frac{1}{k!} \\
&\leq 2+\sum_{k=2}^{n} \frac{1}{2^{k-1}} \leq 2+1 = 3.
\end{aligned}$$

■

Exercise 188.
Show that for any $x \neq y \in \mathbb{R}$, $x, y > 0$, it is the case that

$$\sqrt[n+1]{x^n y} < \frac{y+nx}{n+1}.$$

Solution.
Use the AG-MEAN with $a_i = n$, for $1 \leq i \leq n$ and $a_{n+1} = y$. Then

$$\sqrt[n+1]{x^n y} < \frac{\overbrace{x+x+\ldots+x}^{n \text{ times}}+y}{n+1} = \frac{nx+y}{n+1}.$$

■

Exercise 189.
(a) Show that the sequence

$$a_n = \left(1+\frac{1}{n}\right)^n$$

is an increasing sequence on n.
(b) Show that the sequence

$$b_n = \left(1-\frac{1}{n}\right)^n$$

is an increasing sequence on n.

Solution.
(a) In $\sqrt[n+1]{x^n y} < \frac{y+nx}{n+1}$ set $y=1$ and $x = 1+1/n$. Then

$$\begin{aligned}
\sqrt[n+1]{x^n y} &< \frac{y+nx}{n+1} \Leftrightarrow \\
\sqrt[n+1]{\left(1+\frac{1}{n}\right)^n} &< \frac{1+n(1+1/n)}{n+1} = 1+\frac{1}{n+1} \Leftrightarrow \\
\left(1+\frac{1}{n}\right)^n &< \left(1+\frac{1}{n+1}\right)^{n+1}.
\end{aligned}$$

(a) In $\sqrt[n+1]{x^n y} < \frac{y+nx}{n+1}$ set $y=1$ and $x = 1-1/n$.

$$\begin{aligned}
\sqrt[n+1]{x^n y} &< \frac{y+nx}{n+1} \Leftrightarrow \\
\sqrt[n+1]{\left(1-\frac{1}{n}\right)^n} &< \frac{1+n(1-1/n)}{n+1} = 1-\frac{1}{n+1} \Leftrightarrow \\
\left(1-\frac{1}{n}\right)^n &< \left(1-\frac{1}{n+1}\right)^{n+1}.
\end{aligned}$$

■

Exercise 190.
Show that the sequence

$$c_n = \left(1+\frac{1}{n}\right)^{n+1}$$

is a decreasing sequence on an increasing n.

Solution.
It suffices to show the following

$$c_{n+1} < c_n \Leftrightarrow \left(1+\frac{1}{n+1}\right)^{n+2} < \left(1+\frac{1}{n}\right)^{n+1}.$$

$$\begin{aligned} c_n &= \left(1+\frac{1}{n}\right)^{n+1} \\ &= \left(\frac{n+1}{n}\right)^{n+1} \\ &= \frac{1}{\left(\frac{n}{n+1}\right)^{n+1}} \\ &= \frac{1}{\left(1-\frac{1}{n+1}\right)^{n+1}} \\ &= \frac{1}{b_{n+1}}. \end{aligned}$$

Since from the previous problem b_n increases with n, then c_n decreases. ■

Exercise 191.
Show that the limit

$$2 < \lim_{n\to\infty} a_n < 4.$$

Solution.
Observe

$$a_1 = 2 < a_2 < \ldots < a_n < \ldots$$

and

$$c_1 = 4 > c_2 = 3.375 > \ldots > c_n > \ldots$$

Moreover

$$a_n = (1+1/n)^n < c_n = (1+1/n)^{n+1}.$$

Therefore,

$$2 = a_1 < a_2 < \ldots < a_n < c_n < c_{n-1} < \ldots < c_2 = 3.375 < c_1 = 4.$$

Note that $c_4 = 3.0517578125$ and $c_5 = 2.985984$, and thus $a_n < c_5 < c_4$ and therfore $a_n < 3$. Call the limit of a_n with the letter e.

$$2 < e = \lim_{n\to\infty} a_n < 2.985984.$$

■

Exercise 192.
Show that the limit

$$2 < \lim_{n\to\infty} c_n < 4,$$

or

$$\lim_{n\to\infty} c_n = e.$$

For the definition of e see the previous problem.

Solution.

$$\lim(1+\frac{1}{n})^{n+1} = \lim(1+\frac{1}{n})^n(1+1/n) = e\cdot 1 = e.$$

Moreover from the last equation we have the following

$$\left((1+\frac{1}{n})^n\right)(1+1/n) \geq e.$$

■

Exercise 193.
(a) Show that

$$n! > \left(\frac{n}{e}\right)^n.$$

(b) Show that

$$30! > 10^{30}$$

Solution.
(a) Use induction. For $n = 1$ it is obviously true. For $n = k$ let us suppose that $k! > (k/e)^k$. Then for $n = k+1$ we have the following, after using the inductive hypothesis for $n = k$.

$$\begin{aligned}
(k+1)! &= (k+1)\cdot k! \\
&\geq (k+1)\cdot (k/e)^k \\
&= \frac{k^k}{(k+1)^k}\cdot e\cdot \left(\frac{k+1}{e}\right)^{k+1} \\
&= \frac{e}{(1+\frac{1}{k})^k}\cdot \left(\frac{k+1}{e}\right)^{k+1} \\
&\geq 1\cdot \left(\frac{k+1}{e}\right)^{k+1} \\
&= \left(\frac{k+1}{e}\right)^{k+1},
\end{aligned}$$

and this completes the inductive step.
(b) Since $n! > \left(\frac{n}{e}\right)^n$ substituting 30 for n, and noting $e < 3$, we obtain the following.

$$30! > (30/e)^{30} > (30/3)^{30} = 10^{30}.$$

■

Exercise 194.
Show that

$$n! < e\left(\frac{n+1}{e}\right)^{n+1}.$$

Solution.
Use induction. For $n = 1$ it is obviously true. For $n = k-1$ let us suppose that $(k-1)! < e \cdot (k/e)^k$. Then for $n = k$ we have the following, after using the inductive hypothesis for $n = k-1$.

$$\begin{aligned} k! &= k \cdot (k-1)! \\ &\leq k \cdot e \cdot (k/e)^k \\ &= \frac{k^{k+1}}{(k+1)^{k+1}} \cdot e^2 \cdot \left(\frac{k+1}{e}\right)^{k+1} \\ &= \frac{e^2}{(1+\frac{1}{k})^{k+1}} \cdot \left(\frac{k+1}{e}\right)^{k+1} \\ &\leq e \cdot \left(\frac{k+1}{e}\right)^{k+1} \\ &= e \cdot \left(\frac{k+1}{e}\right)^{k+1}, \end{aligned}$$

utilizing the fact that $c_n = (1+1/n)^{n+1} > e$ for $n = k$, and this completes the inductive step.

■

Exercise 195.
(a) Show that for positive integer $n > 0$, we have the following.

$$\left(1+\frac{1}{n}\right)^n \leq e \Leftrightarrow \ln\left(1+\frac{1}{n}\right) \leq \frac{1}{n}. \tag{4.15}$$

(b) Show that for positive integer $n > 0$, we have the following.

$$\left(1+\frac{1}{n}\right)^{n+1} \geq e \Leftrightarrow \frac{1}{n+1} \leq \ln\left(1+\frac{1}{n}\right). \tag{4.16}$$

(c) Conclude then the following

$$\frac{1}{n+1} < \ln\left(1+\frac{1}{n}\right) < \frac{1}{n}. \tag{4.17}$$

Solution.
(a) Eq. 4.15 was proven ealier as shown below in (c). One can prove Eq. 4.15 using the binomial theorem.

$$(1+\frac{1}{n})^n = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n^k} 1^{n-k} = \sum_{k=0}^{n} \frac{n!}{k!\cdot(n-k)!\cdot n^k} \leq \sum_{k=0}^{n} \frac{n(n-1)\ldots(n-k+1)}{k!\cdot n^k} \leq \sum_{k=0}^{n} \frac{1}{k!} \leq \sum_{k=0}^{\infty} \frac{1}{k!} = e$$

(b) In order to prove

$$\left(1+\frac{1}{n}\right)^{n+1} \geq e \Leftrightarrow (n+1)\ln\left(1+\frac{1}{n}\right) \geq 1.$$

We switch variables $t = 1/n$, where $0 \leq t \leq 1$. Proving this becomes equivalent to proving the following

$$(1+1/t)\ln(1+t) \geq 1 \Leftrightarrow \ln(1+t) \geq \frac{t}{t+1}.$$

It suffices then to show that

$$f(t) = \ln(1+t) - \frac{t}{t+1} \geq 0.$$

We first find $f'(t)$. It is $f'(t) = t/(t+1)^2$ which is ≥ 0 for $t \geq 0$. Thus f is increasing. For $f(0) = 0$ this implies $f(t) \geq f(0) = 0$ and thus $f(t) \geq 0$ as needed. This completes the proof.
(c) We can get the conclusion directly from what was proven in previous problems. From previous problems we have the following.

$$\left(1+\frac{1}{n}\right)^n < e < \left(1+\frac{1}{n}\right)^{n+1}.$$

Take the logarithms base e of all three sides. The result follows

$$\left(1+\frac{1}{n}\right)^n < e \Leftrightarrow \ln\left(1+\frac{1}{n}\right) < \frac{1}{n},$$

$$e < \left(1+\frac{1}{n}\right)^{n+1} \Leftrightarrow 1 < (n+1)\ln\left(1+\frac{1}{n}\right) \Leftrightarrow \frac{1}{n+1} < \ln\left(1+\frac{1}{n}\right).$$

■

Exercise 196.
(a) Show the following.

$$e^x \geq 1+x \quad \forall x \in R. \tag{4.18}$$

(b) Show that for positive integer $n > 0$, we have the following.

$$e^{\frac{1}{n}} \geq \frac{n+1}{n} \geq 1+\frac{1}{n}. \tag{4.19}$$

Then deduce the following.

$$\left(1+\frac{1}{n}\right)^n \leq e \Leftrightarrow \ln\left(1+\frac{1}{n}\right) \leq \frac{1}{n}. \tag{4.20}$$

(c) Show that for positive integer $n > 0$, we have the following.

$$e^{-\frac{1}{n+1}} \geq \frac{n}{n+1} \geq 1-\frac{1}{n+1}. \tag{4.21}$$

Then deduce the following.

$$\left(1+\frac{1}{n}\right)^{n+1} \geq e \Leftrightarrow \frac{1}{n+1} \leq \ln\left(1+\frac{1}{n}\right). \tag{4.22}$$

(d) Conclude then the following

$$\frac{1}{n+1} < \ln\left(1+\frac{1}{n}\right) < \frac{1}{n}. \tag{4.23}$$

Solution.
The previous result without limits.
(a) Use McLaurin's expansion of e^x or consider $f(x) = e^x - 1 - x$. We have $f'(x) = e^x - 1$. It is $f'(x) = 0$ for $x = 0$. Given that $f''(x) = e^x > 0$ for all x, $f(x)$ has a minimum for $x = 0$. Thus $f(x) \geq f(0) = 0$ for all x.
(b) Consider $x = 1/n$ and substitute for x in (a).

$$e^{\frac{1}{n}} \geq \frac{n+1}{n} \geq 1+\frac{1}{n} \Leftrightarrow e \geq \left(1+\frac{1}{n}\right)^n \Leftrightarrow \left(1+\frac{1}{n}\right)^n \leq e \Leftrightarrow \ln\left(1+\frac{1}{n}\right) \leq \frac{1}{n}.$$

(c) Consider $x = -1/(n+1)$ and substitute for x in (a).

$$e^{-\frac{1}{n+1}} \geq 1-\frac{1}{n+1} \geq \frac{n}{n+1} \geq \frac{1}{\left(1+\frac{1}{n}\right)} \Leftrightarrow e \leq \left(1+\frac{1}{n}\right)^{n+1} \Leftrightarrow \frac{1}{n+1} \leq \ln\left(1+\frac{1}{n}\right).$$

(d) From (b) and (c). ■

Exercise 197.

(a) For integer $n > 0$ show that the following holds.

$$\frac{1}{n+1} \leq \ln(1+\frac{1}{n}) \leq \frac{1}{n}$$

(b) For integer $n > 1$ show that the following holds.

$$\frac{1}{n} \leq -\ln(1-\frac{1}{n}) \leq \frac{1}{n-1}$$

Solution.
Both can be proven as a consequence of the Eq.(4.15) to Eq.(4.23) of the previous problem. We provide two different proofs. The second set uses integration and is quite involved. The first set below uses properties of the expansion of e as used in the previous problem(s).
(a) See the derivations of Eq.(4.20) Eq.(4.22) and subsequently Eq.(4.23) in the previous problem.
(b) Consider

$$e^{-\frac{1}{n}} \geq 1-\frac{1}{n}, \Leftrightarrow \frac{1}{e^{\frac{1}{n}}} \geq 1-\frac{1}{n},$$

and raise both sides to the n-th power, and the left-hand side follows.
Likewise consider,

$$e^{\frac{1}{n-1}} \geq 1+\frac{1}{n-1},$$

and raise both sides to the $(n-1)$-st power, and the right-hand side follows.
The more elaborate proof methods follow.
(a) Graph $y = 1/x$ on the Euclidean plane. Consider $x = 1$ and $x = 1+1/n$. The curve $y = 1/x$ between the two x points is bounded below by a rectangle of x side $1/n$ and y side $1/(1+1/n)$. It is also bounded above by a rectangle of x side $1/n$ and y side 1. Then we have the following.

$$\begin{aligned}
\frac{1}{n}\cdot\frac{1}{1+1/n} &\leq \int_1^{1+1/n}\frac{1}{x}dx \leq \frac{1}{n}\cdot 1 \iff \\
\frac{1}{n+1} &\leq \ln(1+\frac{1}{n}) \leq \frac{1}{n} \iff \\
\frac{n}{n+1} &\leq n\cdot\ln(1+\frac{1}{n}) \leq 1 \iff \\
\frac{n}{n+1} &\leq \ln(1+\frac{1}{n})^n \leq 1 \iff
\end{aligned}$$

The second part above proves (a). Furthermore, We take the limit $\lim_{n\to\infty}\ln(1+\frac{1}{n})^n$. We have

$$\lim_{n\to\infty}\ln(1+\frac{1}{n})^n \geq \lim_{n\to\infty}\frac{n}{n+1} = 1$$

and

$$\lim_{n\to\infty}\ln(1+\frac{1}{n})^n \leq \lim_{n\to\infty} 1 = 1$$

Therefore $\lim_{n\to\infty}\ln(1+\frac{1}{n})^n = 1$ or equivalently $\lim_{n\to\infty}(1+\frac{1}{n})^n = e$.
(b) Graph $y = 1/x$ on the Euclidean plane. Consider $x = 1$ and $x = 1-1/n$. The curve $y = 1/x$ between the two

x points is bounded above by a rectangle of x side $1/n$ and y side $1/(1-1/n)$. It is also bounded below by a rectangle of x side $1/n$ and y side 1. Then we have the following.

$$\begin{aligned} 1 \cdot \frac{1}{n} &\leq \int_{1-1/n}^{1} \frac{1}{x} dx \leq \frac{1}{1-1/n} \cdot \frac{1}{n} \iff \\ \frac{1}{n} &\leq -\ln(1-\frac{1}{n}) \leq \frac{1}{n-1} \iff \\ 1 &\leq -n \cdot \ln(1-\frac{1}{n}) \leq \frac{n}{n-1} \iff \end{aligned}$$

The second part above proves (b). Furthermore, We take the limit $\lim_{n\to\infty} -\ln(1-\frac{1}{n})^n$. We have

$$\lim_{n\to\infty} -\ln(1-\frac{1}{n})^n \geq \lim_{n\to\infty} 1 = 1$$

and

$$\lim_{n\to\infty} -\ln(1-\frac{1}{n})^n \leq \lim_{n\to\infty} \frac{n}{n-1} = 1$$

Therefore $\lim_{n\to\infty} -\ln(1-\frac{1}{n})^n = 1$ or equivalently $\lim_{n\to\infty}(1-\frac{1}{n})^n = 1/e$. ■

Exercise 198.
Define the Harmonic series of order n as the sum of the first n terms of the harmonic sequence.

$$H_n = \sum_{i=1}^{n} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}.$$

Show that the infinite harmonic series diverges.

$$H = \lim_{n\to\infty} H_n = \sum_{i=1}^{\infty} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots,$$

and

$$\lim_{n\to\infty} H_n \geq \lim_{n\to\infty} \ln(n+1) = \infty.$$

Solution.
From

$$\frac{1}{n+1} < \ln\left(1 + \frac{1}{n}\right) < \frac{1}{n}.$$

we obtain

$$\frac{1}{n} > \ln\left(\frac{n+1}{n}\right).$$

We substitute $n = 1, 2, \ldots, n$ as follows.

$$\begin{aligned}
\frac{1}{1} &> \ln\left(\frac{2}{1}\right) \\
\frac{1}{2} &> \ln\left(\frac{3}{2}\right) \\
\frac{1}{3} &> \ln\left(\frac{4}{3}\right) \\
&\ldots \\
\frac{1}{n-1} &> \ln\left(\frac{n}{n-1}\right) \\
\frac{1}{n} &> \ln\left(\frac{n+1}{n}\right)
\end{aligned}$$

We add all those inequalities. We get the following.

$$H_n = \sum_{i=1}^{n} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \geq \ln(n+1).$$

Therefore

$$\lim_{n\to\infty} H_n \geq \lim_{n\to\infty} \ln(n+1) = \infty.$$

■

Exercise 199.

Define the Harmonic series of order n as the sum of the first n terms of the harmonic sequence.

$$H_n = \sum_{i=1}^{n} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}.$$

Let

$$H = \lim_{n\to\infty} H_n = \sum_{i=1}^{\infty} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots,$$

and then show that the following.

$$\lim_{n\to\infty} H_n < \lim_{n\to\infty} \ln(n) + 1 = \infty.$$

Solution.
From

$$\frac{1}{n+1} < \ln\left(1 + \frac{1}{n}\right) < \frac{1}{n}.$$

we obtain

$$\ln\left(\frac{n+1}{n}\right) > \frac{1}{n+1}.$$

We then substitute $n = 2, \ldots, n$ as follows.

$$\begin{aligned}
\ln\left(\frac{2}{1}\right) &> \frac{1}{2} \\
\ln\left(\frac{3}{2}\right) &> \frac{1}{3} \\
\ln\left(\frac{4}{3}\right) &> \frac{1}{4} \\
&\ldots \\
\ln\left(\frac{n-1}{n-2}\right) &> \frac{1}{n-1} \\
\ln\left(\frac{n}{n-1}\right) &> \frac{1}{n}
\end{aligned}$$

We add all those inequalities. We get the following.

$$\ln \frac{2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n}{1 \cdot 2 \cdot \ldots \cdot (n-2) \cdot (n-1)} > \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} + \frac{1}{n} = H_n - 1 \Leftrightarrow \ln(n) > H_n - 1 \Leftrightarrow \ln(n) + 1 > H_n.$$

Therefore

$$\lim_{n\to\infty} H_n < \lim_{n\to\infty} \ln(n) + 1 = \infty.$$

■

4.8 Exponential and logarithmic inequalities

The primary source for the listed inequalities is [1], [29], [2] , [21].

Exercise 200.
(a) Show the following for $x \in \mathbb{R}$.

$$x < e^x - 1. \tag{4.24}$$

(b) Show the following for $x \in \mathbb{R}$, with $x \in (0,1)$.

$$e^x - 1 < \frac{x}{1-x} \Leftrightarrow e^x < \frac{1}{1-x} \tag{4.25}$$

(c) Show the following for $x \in \mathbb{R}$, with $x \in (-\infty, 0)$.

$$e^x - 1 < \frac{x}{1-x} \Leftrightarrow e^x < \frac{1}{1-x} \tag{4.26}$$

(d) Show the following for all $x \in \mathbb{R}$ with

$$1 - x < \exp(-x). \tag{4.27}$$

(e) Show the following for all $x \in \mathbb{R}$ with $x \in (-\infty, 0)$.

$$\exp(-\frac{x}{1-x}) < 1 - x \Leftrightarrow \frac{1}{1-x} < \exp(\frac{x}{1-x}). \tag{4.28}$$

(f) Show the following for all $x \in \mathbb{R}$ with $x \in (0,1)$.

$$\exp(-\frac{x}{1-x}) < 1 - x \Leftrightarrow \frac{1}{1-x} < \exp(\frac{x}{1-x}). \tag{4.29}$$

Solution.
(a) The $x < e^x - 1$ has already been proved by way of showing that $f(x) = e^x - x - 1 > 0$ for all $x \in \mathbb{R}$. Omitted.
(b) We want to show that $f(x) = e^x - 1 - \frac{x}{1-x} = e^x - \frac{1}{1-x} < 0$. Note that the function $f(x)$ as defined, has a maximum for $x = 0$ since $f'(x) = e^x - 1/(1-x)^2$ with $f'(0) = 0$, and $f''(x) = e^x - 2/(1-x)^3$ with $f''(0) = -1 < 0$. For $0 < x < 1$ one can show $e^x < 1/(1-x)$ by way of series expansion. The expansion of

$$e^x = \sum_i x^i / i! \leq \sum_i x^i = 1/(1-x), \quad (0 < x < 1)$$

for all $i \geq 1$.
(c) For the case $x < 0$ change variables first: use $-t = x < 0$ i.e. $t > 0$ to obtain

$$e^x < \frac{1}{1-x} \Leftrightarrow \frac{1}{e^t} < \frac{1}{1+t} \Leftrightarrow e^t > 1 + t.$$

The latter is inequality (a).
(d) Set $x = -x$ in (a).
(e) Change variables $t = -x$ and since $x < 0$ we have $t > 0$ subsequently deriving the following.

$$\exp(-\frac{x}{1-x}) < 1 - x \Leftrightarrow \exp(\frac{t}{1+t}) < 1 + t \Leftrightarrow \frac{t}{1+t} < \ln(1+t).$$

We examine $g(t) = \frac{t}{1+t} - \ln(1+t)$ for $t > 0$. We need to show $g(t) < 0$. Consider $g'(t) = 1/(1+t)^2 - 1/(1+t)$. It is $g'(t) < 0$ for all $t > 0$. Thus $g(t)$ is decreasing in $t > 0$. Therefore $g(t) < g(0) = 0$ and the result follows.
(f) Change variables: $t = x/(1-x)$ i.e. $x = t/(1+t)$. Since $x > 0$ and $1 - x > 0$ we have $t > 0$ as well. Then

$$\exp(-\frac{x}{1-x}) < 1 - x \Leftrightarrow \frac{1}{1-x} < \exp(\frac{x}{1-x}) \Leftrightarrow 1 + t < \exp(t) = e^t.$$

The latter is part (a). ■

Exercise 201.
(a) Show the following for $x \in \mathbb{R}$.

$$x < e^x - 1. \tag{4.30}$$

(b) Show that for any $x \in \mathbb{R}$ we have the following.

$$e^{-x} > 1 - x \Leftrightarrow 1 - e^{-x} < x. \tag{4.31}$$

(c) Show that for any $x \in \mathbb{R}$ with $x > -1$ we have the following.

$$e^{-x} < \frac{1}{1+x} \Leftrightarrow 1 - e^{-x} > \frac{x}{1+x}. \tag{4.32}$$

(d) Show that for any $x \in \mathbb{R}$ with $0 > x > -1$ we have the following.

$$1 + x > \exp\left(\frac{x}{1+x}\right). \tag{4.33}$$

(e) Show that for any $x \in \mathbb{R}$ with $x > 0$ we have the following.

$$1 + x > \exp\left(\frac{x}{1+x}\right). \tag{4.34}$$

(f) Show that for any $0 < x \leq 1.5936$ we have the following.

$$e^{-x} < 1 - (x/2). \tag{4.35}$$

Solution.
(a) It is Eq.(4.24). Omitted.
(b) Set $x = -x$ in (a). Omitted.
(c) This is similar to Eq.(4.25) and Eq.(4.26) for $x \in (-\infty, 1)$ there. This one uses $-x$ for x there and $x \in (-1, \infty)$ instead.
(d) Similar to Eq.(4.27) for $x = -x$.
(e) Similar to Eq.(4.28) for $x = -x$.
(f) Consider $f(x) = 1 - x/2 - e^{-x}$. We are going to show that $f(x) > 0$ in the desired interval $0 < x \leq 1.5936$. Note $f'(x) = e^{-x} - 0.5$. Setting $f'(x) =$ and solving for x we get $x = \ln 2 \approx 0.69314$. We then find $f''(x) = -e^{-x} < 0$. This means $x = \ln 2$ is a maximum, and f is a concave function, in fact stricly concave. Note that $h(0) = 0$, $h(\ln 2) = 0.5 - 0.5 \ln 2 \approx 0.15342$. Furthermore $h(1.593635) \approx 0$. $f'(x) > 0$ for $0 < x < \ln 2$ as $e^{-x} - 0.5 > 0$ and decreasing for $x > \ln 2$. Because of its concavity $f(x) > 0$ for $0 < x \leq 1.5936 < 1.593635$, and the result follows. ■

Exercise 202.
(a) For every $x > -1$ we have

$$\ln(1+x) \leq x. \tag{4.36}$$

(b) For every $x \geq 0$ we have

$$\ln(1+x) \leq x/(1+x). \tag{4.37}$$

Solution.
(a) Take logarithms of both sides of Eq.(4.32).
(b) Take logarithms of both sides of Eq.(4.34). ■

Exercise 203.
(a) Show that for every $x>0, y>0$ we have the following.

$$e^x > (1+\frac{x}{y})^y. \tag{4.38}$$

(b) Show that for every $x>0, y>0$ we have the following.

$$(1+\frac{x}{y})^y > e^{\frac{xy}{x+y}}. \tag{4.39}$$

Solution.
(a) Since

$$e^x > (1+\frac{x}{y})^y \Leftrightarrow x > y\ln(1+\frac{x}{y}) \Leftrightarrow x/y > \ln(1+\frac{x}{y}),$$

it suffice to show the latter. Consider $t=x/y>0$ since $x,y>0$, and

$$f(t) = t - \ln(1+t).$$

We have $f'(t) = 1-1/(1+t)$. It is $f'(t)>0$ for all $t>0$, Evaluate $f(0)$ which is $f(0)=0$. Thus $f(t)>f(0)=0$ for all t.
(b) Since $x>0$, $y>0$ and the following are are equivalent with the last inequality derived with a change of variable $t=x/y>0$,

$$(1+\frac{x}{y})^y > e^{\frac{xy}{x+y}} \Leftrightarrow y\ln(1+\frac{x}{y}) > \frac{xy}{x+y} \Leftrightarrow \ln(1+\frac{x}{y}) > \frac{x}{x+y} \Leftrightarrow \ln(1+t) > \frac{t}{1+t},$$

it suffices to show that $f(t) = \ln(1+t) - \frac{t}{1+t} > 0$ for all $t>0$. We have $f'(t) = 1/(1+t) - 1/(1+t)^2 > 0$ and thus $f(t)>f(0)=0$, as needed. ■

Exercise 204.
For every $x > 0$ show the following.

$$\ln(1+x) \geq \frac{x}{1+x/2}. \tag{4.40}$$

Solution.
Consider $f(x) = \ln(1+x) - \frac{x}{1+x/2}$. Then $f'(x) = \frac{x^2}{(1+x)(2+x)^2}$. For $x > -1$ we have $f'(x) > 0$, and thus $f(x)$ is increasing. $f(x) > f(0) = 0$ and the result follows.
One could also prove a slighly better bound for $x > -1$

$$\frac{1}{1+x/2} < \ln(1+x)/x,$$

utilizing

$$f(x) = \ln(1+x) - \frac{x}{1+x/2}$$

with the case $x < 0$, leading to an $f(x) \leq f(0) = 0$ and then from $\ln(1+x) \leq \frac{x}{1+x/2}$ concluding by mulitplying with a negative x that $\frac{1}{1+x/2} < \ln(1+x)/x$.

■

Exercise 205.

See [1].
(a) For every $0 < t < 0.45$ we have the following.

$$\ln(1+t) > t - t^2/2 + t^3/4.$$

(b) For every $0 < t < 0.68$ we have the following.

$$\ln(1-t) > -t - t^2.$$

(c) For every $0 < t < 0.431$ we have the following.

$$\ln(1-t) > -t - t^2 - t^3/3.$$

Show the results as claimed.
In the second inequality a 0.69 of [1] has been modified to 0.68.

Solution.
Omitted. See [1].
(a) Consider function

$$f(t) = \ln(1+t) - t + t^2/2 - t^3/4,$$

and determine its minimum (or infimum). We have

$$f'(t) = \frac{t^2(1-3t)}{4(1+t)}.$$

Therefore $f'(t) =$ for $t = 0$ and $t = 1/3$. Moreover for $t = -1$ is undetermined. Only $t = 0$ and $t = 1/3$ are inside or the neighborhood of $(0, 0.45)$. We observe $f'(t) > 0$ for $0 < t < 1/3$. We observe $f'(t) < 0$ for $1/3 < t < 0.45$. Thus a max occurs at $t = 1/3$. Furthermore at $f(0) = 0$ and $f(0.45) = 0.0000323$. Therefore $f(t) > 0$ for $f(0) = 0$ being an infimum.
(b) Similarly as before

$$f(t) = \ln(1-t) + t + t^2,$$

and

$$f'(t) = \frac{t(1-2t)}{(1-t)}.$$

For $t = 1$ (out of range) $f'(t)$ is undefined, but $t = 0$ and $t = 1/2$ give $f'(t) = 0$. For $0 < t < 1/2$ we have $f'(t) > 0$ thus increasing; for $1/2 < t < 0.68$ we have $f'(t) < 0$ decreasing. The max occurs at $t = 1/2$. We are interested in the min. For $t = 0$ we have $f(0) = 0$ and $t = 0.68$ we have $t = 0.0029657$. Thus 0 is an infimum as $t = 0$ is out of range. Therefore $f(t) > f(0)$ as needed.
(c) Similarly as before

$$f(t) = \ln(1-t) + t + t^2 + t^3/3,$$

and

$$f'(t) = \frac{t(1-t-t^2)}{(1-t)}.$$

Function $f'(t) > 0$ for all $0 < t < 0.431$ and thus $f(t) > f(0) = 0$ in this range. ■

Exercise 206.
(a) For every $x > -1$ and $x \neq 0$ we have the following.

$$\frac{x}{1+x} < \ln(1+x) < x. \tag{4.41}$$

(b) For every $x > 0$ we have the following.

$$\ln(x) \leq x - 1. \tag{4.42}$$

(c) For every $x < 1$ and $x \neq 0$ we have the following.

$$x < -\ln(1-x) < \frac{x}{1-x}. \tag{4.43}$$

Solution.
(a)
Left-hand side first. One can claim that

$$\frac{x}{1+x} < \ln(1+x)$$

can be obtained from Eq.(4.33) and Eq.(4.34) for the $-1 > x > 0$ and $x > 0$ cases and then unifying them.
A direct (standalone) proof is to start with

$$f(x) = \ln(1+x) - \frac{x}{1+x}$$

and prove $f(x) > 0$. Since $f'(x) = x/(1+x)^2$, for $x > 0$, $f(x)$ is non-decreasing and thus $f(x) \geq f(0) = 0$ and the result follows. For $-1 < x < 0$, we have $f'(x) \leq 0$ and thus $f(x)$ is non-increasing and thus $f(x) \geq f(0) = 0$ and the result follows.
Right-hand side next. The right hand-side is by (4.47). One can alternatively claim that

$$\ln(1+x) < x$$

can be obtained by showing $f(x) = \ln(1+x) - x$, has $f'(x) = 1/(1+x) - 1 = 0$ for $x = 0$ and $f''(x) = -1/(1+x)^2 < 0$ for all relevant values of x (concave function), indicating a maximum for $x = 0$, thus indicating $f(x) < f(0)$, $x \neq 0$.
(b) It follows from (4.49), Alternatively, we can use $f(x) = x - 1 - \ln(x)$ and show $x = 1$ is a minimum of $f(x)$ impying $f(x) \geq f(1) = 0$.
(c) The third part inequality's left-hand size follows from (4.48). The right hand side follows by utilizing the method used for the first part's left hand-side. Alternatively, the left-hand side for $x \in (0,1)$ and $x \in (-\infty, 0)$ can be obtained from Eq.(4.25) and Eq.(4.26), and then unifying the cases. The right-hand side for $x \in (-\infty, 0)$ and $x \in (0,1)$ can be obtained from Eq.(4.28) and Eq.(4.29), and then unifying the cases. ■

Exercise 207.
For every $x > -1$ and $x \neq 0$ the following is true.

$$\frac{x}{1+x} < \ln(1+x) \leq \frac{x(6+x)}{6+4x} < x. \tag{4.44}$$

Solution.
The left-hand side is by the previous problem. Utilize then

$$x^4 + 4y^4 = (x^2 + 2xy + y^2)(x^2 - 2xy + 2y^2).$$

■

Exercise 208.
Show the following. For every $x \in \mathbb{R}$ with $x \in (0, \infty)$ we have the following.

$$\frac{2x}{2+x} \leq \ln(1+x) \leq \frac{x(x+2)}{2 \cdot (x+1)}.$$

For every $x \in \mathbb{R}$ with $x \in (-1, 0)$ we have the following.

$$\frac{2x}{2+x} \geq \ln(1+x) \geq \frac{x(x+2)}{2 \cdot (x+1)}$$

Solution.
Omitted. ■

Exercise 209.
For every $0 < x < 0.5828$ we have the following.

$$|\ln(1-x)| < 1.5x.$$

Solution.
Omitted. ■

4.9 Series

Exercise 210.
Show that $\sum_{i=1}^{n} i^3 \leq n^4$.

Solution.

$$\sum_{i=1}^{n} i^3 = 1^3 + 2^3 + 3^3 + \ldots + i^3 + \ldots + (n-1)^3 + n^3.$$

Since i runs from 1 to n we have, using the examples before,

$$\begin{aligned}
1 \leq i \leq n : 1 \leq n \Rightarrow & \quad 1^3 \leq n^3 \\
\ldots \Rightarrow & \quad \ldots \\
i \leq n \Rightarrow & \quad i^3 \leq n^3 \\
\ldots \Rightarrow & \quad \ldots \\
n \leq n \Rightarrow & \quad n^3 \leq n^3 \\
\textbf{Add both sides of inequalities} & \quad 1^3 + 2^3 + \ldots + i^3 + \ldots + n^3 \leq n^3 + n^3 + \ldots + n^3 + \ldots + n^3 \\
& \quad \sum_{i=1}^{n} i^3 \leq n \cdot n^3 = n^4
\end{aligned}$$

■

Exercise 211.
Show that for any $n \geq 2$, $\sum_{i=1}^{n} i \leq 3n^2/4$.
Show that for any $x \geq 3$, $\sum_{i=0}^{n-1} x^i \leq x^n/2$.

Solution.
Omitted. Practice makes perfect. ■

Exercise 212.
(a) Show the following for any $x \geq 2$. See [30].

$$\ln(x) = \sum_{i=1}^{\infty} \frac{1}{i} \cdot \left(\frac{x-1}{x}\right)^i = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \dots \tag{4.45}$$

(b) For $|x-1| \leq 1$ and $x \neq 0$, we have as follows. See [30].

$$\ln(x) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{1}{i} \cdot (x-i)^i = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \dots \tag{4.46}$$

Solution.
(a) Taylor series expansion of $\ln(x)$ at $t = 1$.
(b) Substitute x for $(x-1)/x$ in (a). Then the series of the next problem part (b) results. Subsequently replace $1-x$ for x. ■

Exercise 213.
(a) The McLaurin series for $\ln(1+x)$ is as follows. It is also known as Mercator-Newton series. The series converges for $|x|<1$ and $x=1$. Show the claims.

$$\ln(1+x)=\sum_{i=1}^{\infty}(-1)^{i+1}\frac{x^i}{i}=x-\frac{x^2}{2}+\frac{x^3}{3}-\dots \tag{4.47}$$

(b) The McLaurin series for $\ln(1-x)$ is as follows. The series converges for $|x|<1$ and $x=-1$.

$$\ln(1-x)=-\sum_{i=1}^{\infty}\frac{x^i}{i}.=-x-\frac{x^2}{2}-\frac{x^3}{3}-\dots \tag{4.48}$$

Solution.
(a) Consider the Taylor expansion of $\ln(x)$ at $t=1$ in Eq.(4.45) of the previous problem, and then substitute $1+x$ for x.
(b) Substitute $-x$ for x in the previous one. ■

Exercise 214.
(a) As a consequence of Eq. (4.47), the following hold for $-1 < x \leq 1$.

$$\ln(1+x) \leq x - x^2/2 + x^3/3, \tag{4.49}$$

(b) As a consequence of Eq. (4.47), the following hold for $-1 < x \leq 1$.

$$\ln(1+x) \geq x - x^2/2. \tag{4.50}$$

(c) Extend (a) for $x > 0$.
(d) Extend (b) for $x > 0$.

Solution.
(a) Omitted.
(b) Omitted.
(c) Consider,

$$f(x) = \ln(1+x) - x + x^2/2 - x^3/3,$$

and show $f(x)$ is decreasing with maximum for $x = 0$.
(d)

$$f(x) = \ln(1+x) - x + x^2/2,$$

and show $f(x)$ is increasing with minimum for $x = 0$.

■

Exercise 215.
As a consequence of Eq. (4.48), the following hold for $-1 \leq x < 1$.

$$\ln(1-x) \leq \sum_{i=1}^{t} (-1)\frac{x^i}{i}, \tag{4.51}$$

and

$$\ln(1-x) \geq \sum_{i=1}^{t} (-1)\frac{x^i}{i} - \frac{x^t}{t}. \tag{4.52}$$

Solution.
Omitted. ■

Exercise 216.
For every $2 \geq x > 0$ we have the following.

$$\ln(x) \geq \sum_{i=1}^{\infty} (-1)^{i+1} (x-1)^i / i \tag{4.53}$$

Solution.
Consider $1/(1-(1-x)$. It converges for $|1-x| < 1$ that is $1-x < 1$ which implies $x > 0$ and $-1 < 1-x$ which implies $x < 2$. Expanding $1/(1-(1-x)$ in powers of $1-x)$ and integrating one can find the Taylor expanstion at $t = 1$. ■

Exercise 217.
For every $0 \leq \delta \leq 1$ we have the following.

$$(1+\delta)\ln(1+\delta) \geq \delta + \delta^2/3.$$

Solution.
For $0 \leq x \leq 1$,

$$\begin{aligned} \ln(1+x) &= x - x^2/2 + x^3/3 - x^4/4 + \dots \\ (1+x)\ln(1+x) &= x - x^2/2 + x^3/3 - x^4/4 + \dots \\ &+ x^2 - x^3/2 + x^4/3 - x^5/4 + \dots \\ (1+x)\ln(1+x) &\geq x + x^2/2 - x^3/6 \end{aligned}$$

We can then conclude since $0 \leq x \leq 1$ that

$$(1+x)\ln(1+x) \geq x + x^2/2 - x^3/6 \geq x + x^2/2 - x^2/6 \geq x + x^2/3.$$

Set $x = \delta$ and the result follows. ■

Exercise 218.
For every $2e-1<\delta$ we have the following.

$$(1+\delta)\ln(1+\delta)\geq(2e)^{2e}.$$

Solution.
Omitted. ■

Exercise 219.
For every $0 \leq \delta < 1$ we have the following.

$$(1-\delta)\ln(1-\delta) \geq -\delta + \delta^2/2.$$

Solution.
For $0 \leq x < 1$,

$$\begin{aligned}
\ln(1-x) &= -\sum_{i=1}^{\infty} x^i/i \\
&= -x - x^2/2 - x^3/3 - x^4/4 - \dots \\
(1-x)\ln(1-x) &= -x - x^2/2 - x^3/3 - x^4/4 - \dots \\
&+ \quad +x^2 + x^3/2 + x^4/3 + x^5/4 + \dots \\
(1-x)\ln(1-x) &= -x + x^2/2 + x^2/6 + \dots
\end{aligned}$$

The missing terms in the last form of the equation are positive. We can then conclude

$$(1-x)\ln(1-x) \geq -x + x^2/2 + x^2/6 \geq -x + x^2/2$$

■

Exercise 220.
The following holds.

$$\ln(n+1)-\ln(n-1)=2\left(\frac{1}{n}+\frac{1}{3n^3}+\frac{1}{5n^5}+\dots\right).$$

Solution.
Combine Eq.(4.47) and Eq.(4.48) and then substitute $1/n$ for x in $\ln\frac{1+x}{1-x}$. ■

Exercise 221.
The following holds.

$$\exp(-\frac{1}{n}) = \sum_{i=1}^{\infty}(-1)^i \frac{1}{i! \cdot n^i} = 1 - \frac{1}{n} + \frac{1}{2! \cdot n^2} - \frac{1}{3! \cdot n^3} \geq 1 - \frac{1}{n}.$$

Solution.
McLaurin expansion of e^x where $x = -1/n$. Omitted. ■

4.10 Entropy and probability

Exercise 222.

(a) Find the extrema of $F(x) = x\lg(x) + (n-x)\lg(n-x)$, $0 < x < n$ for $n > 0$.
(b) Find the minimum of the entropy function $f(x) = x\lg x + (1-x)\lg(1-x)$. $0 < x < 1$ and relate it to part (a).

Solution. For $0 < x < n$ we use $y = x/n$ i.e. $x = yn$, and then $0 < y < 1$. Then

$$\begin{aligned} F(x) &= x\lg(x) + (n-x)\lg(n-x) \\ F(yn) &= yn\lg(yn) + (n-yn)\lg(n-yn) = yn\lg y + yn\lg n + n(1-y)\lg(1-y) + n(1-y)\lg n \\ &= yn\lg y + n(1-y)\lg(1-y) + n\lg n \\ &= nG(y) + n\lg n \end{aligned}$$

In the remainder we study $G(y)$ with $0 < y < 1$.

$$G(y) = y\lg(y) + (1-y)\lg(1-y).$$

But

$$G(y) = -f(y),$$

where $f(y)$ is the function of the previous problem (it does not matter that we called the function there $f(x)$, the alias x or y does not matter). The maximum of $f(y)$ at $y = 1/2$ (previous problem) translates into a minimum for $G(y)$ at $y = 1/2$ and a minimum for $F(yn)$ and $y = 1/2$ or a minimum for $F(x)$ at $x = yn = n/2$.

If the previous problem was not there, and you do not like change of bases then we proceed as follows.

We first find $f'(y)$.

$$\begin{aligned} f'(y) &= \frac{\ln y + y \cdot 1/y - \ln(1-y) + (1-y)(-1)/(1+y)}{\ln 2} \\ &= \frac{\ln y - \ln(1-y)}{\ln 2}. \end{aligned}$$

$f'(y) = 0$ for $y = 1-y$ i.e. $y = 1/2$. Then we find the second derivative of f.

$$f''(y) = \frac{1}{\ln 2}\left(\frac{1}{y} + \frac{1}{1-y}\right).$$

Then we conclude $f''(1/2) = 4/\ln 2 > 0$. Thus f has a minimum for $y = 1/2$ and so does F at $x = yn = n/2$. ■

Exercise 223.

(Some parts are from a previous problem.)
(a) Find the extrema of $F(x) = x\lg(x) + (n-x)\lg(n-x)$, $0 < x < n$.
Is $F(n/4) \geq n\lg n - n$? Is $F(n/2) \geq n\lg n - n$?
(b) Analyze then entropy function $f(x) = -x\lg x - (1-x)\lg(1-x)$, with $0 < x < 1$, by analysing $g(x) = x\lg x + (1-x)\lg(1-x)$.

Solution. This is primarily Calculus 1. This problem reviews topics related to finding the extrema (min or max) of functions. One needs it to prove the best case running time of QuickSort. Consider.

$$f(x) = -x\lg x - (1-x)\lg(1-x)$$

and $g(x) = -f(x)$ i.e.

$$g(x) = x\lg x + (1-x)\lg(1-x).$$

If $g(x)$ has a minimum or maximum then its negative i.e. $f(x)$ will have a maximum or a minimum respectively.
STEP 1. (Symmetry over $x = 1/2$). The first observation is that $f(x), g(x)$ are symmetric around $x = 1/2$. Thus $f(0) = f(1)$ if they were defined and in general $f(x) = f(1-x)$.
Step 2. (Behavior at $x = 0$. What is the limit $\lim_{x\to 0} g(x)$?

$$\lim_{x\to 0} g(x) = \lim_{x\to 0}(x\lg x + (1-x)\lg(1-x)) = \lim_{x\to 0}(x\lg x + (1-0)\lg(1-0)) = \lim_{x\to 0} x\lg x.$$

Step 3. (Change variables $y = 1/x$.) At this point we change variables $y = 1/x$ or $x = 1/y$. For $x \to 0$, or $1/y \to 0$, we have $y \to \infty$. Therefore

$$\lim_{x\to 0} g(x) = \lim_{x\to 0} x\lg x = \lim_{y\to\infty}(1/y)\lg(1/y) = \lim_{y\to\infty} \frac{-\lg y}{y} \stackrel{L'Hopital}{=} \lim_{y\to\infty} \frac{-(\lg y)'}{(y)'} = \lim_{y\to\infty} \frac{-1/y}{1} = 0^-$$

Thus the limit $\lim_{x\to 0} g(x) = 0$ from the left side (from the negative) and we represent this by having a minus as a superscript to the 0. It also suggests that $\lim_{x\to 1} g(x) = 0^-$ as well by STEP 1. Therefore for $f(x) = -g(x)$ we have $\lim_{x\to 1} f(x) = 0^+ = \lim_{x\to 0} f(x)$.

STEP 4. (First Derivative of g) What is $g'(x)$?
Simple calculations reveal that

$$g'(x) = \lg x - \lg(1-x)$$

The first derivative is 0 for $\lg x = \lg(1-x)$ i.e. for $x = 1/2$. Therefore at $x = 1/2$ we have a minimum or maximum (most likely).
STEP 5. (Second Derivative of g) To determine the type of the extreme (minimum or maximum), we examine the second derivative for $x = 1/2$ and then we obtain that

$$g''(x) = 1/x + 1/(1-x), \text{ and thus } g''(1/2) = 4 > 0.$$

Therefore $g(x)$ has a minimum at $x = 1/2$ and symmetrically $f(x)$ has a maximum. It is easy to find that $g(1/2) = -1$, and thus $f(1/2) = 1$.
STEP 6 ($F(x)$). Use $y = x/n$ in $F(x)$, $0 < x < n$. Then, $x = yn$ with $0 < y < 1$. Then

$$\begin{aligned} F(x) &= x\lg(x) + (n-x)\lg(n-x) \\ F(yn) &= yn\lg(yn) + (n-yn)\lg(n-yn) = yn\lg y + yn\lg n + n(1-y)\lg(1-y) + n(1-y)\lg n \\ &= yn\lg y + n(1-y)\lg(1-y) + n\lg n \\ H(y) &= n(y\lg y + (1-y)\lg(1-y)) + n\lg n = ng(y) + n\lg n \end{aligned}$$

Function $H(y)$ has a constant term (not dependent on y) $n\lg n$ plus $ng(y)$. Therefore it retains a minimum at $y = 1/2$ i.e. $x = yn = n/2$ and $g(1/2) = -1$ by Step 5.

STEP 7 (Conclusion). As far as $F(n/4)$ is concerned, given that $g(x)$ has a minimum for $g(1/2)$ we have that $g(1/4) \geq g(1/2)$ and thus for $x = yn$, i.e. $y = 1/4$ we have that

$$F(n/4) = n\,(g(1/4)) + n\lg n \geq ng(1/2) + n\lg n \geq n\lg n - n.$$

■

4.11 Maxima and Minima

Exercise 224.

A different from. Find the maximum of the entropy function $f(x) = -x\lg x - (1-x)\lg(1-x)$. $0 < x < 1$.

Solution.
We are using logarithms base two. Note that Consider $2^y = e$ i.e. $y = \lg e$ by definition. Taking ln of both sides we hage $y \ln 2 = 1$. Thus $\lg e = y = 1/\ln 2$. It is easy with a calculator to establish that $\lg e > 1$ and in fact and $\lg e \approx 1.442$. Consider now,

$$2^{\lg x} = e^{\ln x} = x.$$

Take lg of both sides. Therefore $\lg x = \ln x \lg e$. We use it to convert $f(x)$ into

$$f(x) = -x\lg x - (1-x)\lg(1-x) = -x\ln x \lg e - (1-x)\ln(1-x)\lg e = g(x)\lg e$$

where $g(x)$ is as follows.

$$g(x) = -x\ln x - (1-x)\ln(1-x).$$

Note that $f^{'}(x) = g^{'}(x)\lg e$. Find the first derivative of $g(x)$.

$$g^{'}(x) = -\ln x - x/x + \ln(1-x) - (1-x)(-1)/(1-x) = -\ln x + \ln(1-x)$$

Setting $g^{'}(x) = 0$ and solving for x we get $x = 1/2$. This is also the case for $f^{'}(x)$. We note that

$$f(1/2) = 1,$$

and that's the minimum value. For this to be the case we should have confirmed earlier that $f''(x) < 0$ or equivalently $g''(x) < 0$. Note that $f^{''}(x) = g^{''}(x)(\lg e)^2$, and the last term is always positive. The monotonicity of $f^{''}(x)$ is the monotonicity of $g^{''}(x)$. Needless

$$g^{''}(x) = -1/x - 1/(1-x)$$

and thus $g^{''}(1/2) = -4 < 0$. This means that $f(x)$ and also $g(x)$ has a maximum at $x = 1/2$. It is not difficult to establish that $f(1/2) = 1$. ■

Exercise 225.

What is the maximum of $f(t) = t^2 + (n-1-t)^2$ for t in the interval $[0, n-1]$?

Solution.
Find the first derivative $f'(t)$, equate it to zero and solve for the indeterminate t to determine the critical point. $f'(t) = 2t + 2(n-1-t)(-1) \Rightarrow f'(t) = 0 \Rightarrow t = (n-1)/2$. The second derivative $f''(t) = 2 + 2(-1)(-1) = 4 > 0$. The function has a minimum at $t = (n-1)/2$. We are not interested in its minimum but rather in its maximum. It is a parabola and thus it is symmetric in the range $[0, n-1]$. The maximum would be at the extreme $t = 0$ or $t = n-1$ or there might be two maxima at $t = 0, t = n-1$. Indeed

$$f(0) = f(n-1) = 0^2 + (n-1-0)^2 = (n-1)^2 + (n-1-(n-1))^2 = (n-1)^2.$$

■

Exercise 226.
What is the minimum of $f(t) = t\lg(t) + (n-1-t)\lg(n-1-t)$ for t in the interval $[0, n-1]$?

Solution.
Find the first derivative $f'(t)$.

$$f'(t) = 1 \cdot \lg t + t \cdot (1/t) + (-1)\lg(n-1-t) + (n-1-t)(-1/(n-1-t)) = \lg(t) - \lg(n-1-t) \Rightarrow$$

$$f'(t) = \lg(t/(n-1-t)).$$

Solving for t in $f'(t) = 0$ relates to $\lg(t) - \lg(n-1-t) = 0$ i.e. $t = n-1-t \Rightarrow t = (n-1)/2$. Finding the second derivative of $f(t)$ i.e. the derivative of $\lg(t) - \lg(n-1-t)$ gives

$$f''(t) = (\lg(t) - \lg(n-1-t))' = 1/t - (-1)/(n-1-t) = 1/t + 1/(n-1-t) > 0$$

Thus the $t = (n-1)/2$ is a minimum. As a conclusion

$$f(\frac{n-1}{2}) = (n-1)\lg((n-1)/2) = (n-1)\lg(n-1) - (n-1).$$

■

Exercise 227.

You are given a random vector $a = (a_1, \ldots a_n)$, where a_i is equally likely and independently to be 0 or 1, i.e. $Pr(a_i = 1) = Pr(a_i = 0) = 1/2$. Answer the following questions.
(a) **(Warmup)** What's the probability that a is the all zero vector ?
(b) Suppose that a, b are two 0-1 vectors of length n whose components were chosen uniformly at random as discussed previously. What is the expected value of the inner product $a \cdot b = \sum_{i=1}^{n} a_i b_i$? Explain.
(c) Let d be a vector of integers mod p (i.e. elements of d are $0, \ldots p-1$), where p is a prime. Let a be a random vector of 0-1's chosen as before. What is an upper bound on the probability that $\sum d_i a_i \equiv 0 \bmod p$? Explain.

Solution.
(a) $1/2^n$.
(b) $n/4$.

$$E[c] = E[ab] = E[\sum_i a_i b_i] = \sum_i E[a_i b_i]$$

$a_i b_i$ is 0 with probability $3/4$ and 1 with probability $1/4$ (when both $a_i = b_i = 1$).

$$E[a_i b_i] = 0(3/4) + 1(1/4) = 1/4$$

Therefore

$$E[c] = E[ab] = E[\sum_i a_i b_i] = \sum_i E[a_i b_i] = n(1/4)$$

(c) The vector d is given (and is not necessarily random). Assume $d \neq 0$, i.e. at least one component of the vector is non-zero since otherwise the problem is trivial. $\sum d_i a_i \equiv 0 \bmod p$.

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \bmod p \\ d_1 a_1 + d_2 a_2 + \ldots + d_n a_n &\equiv 0 \bmod p \\ d_1 a_1 &\equiv (-d_2 a_2 - \ldots - d_n a_n) \bmod p \\ d_1 a_1 &\equiv Z \bmod p \end{aligned}$$

where $Z \equiv (-d_2 a_2 - \ldots - d_n a_n) \bmod p$. Then,

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \bmod p \\ d_1 a_1 &\equiv Z \bmod p \\ a_1 &\equiv (d_1)^{-1} Z \bmod p \\ a_1 &\equiv B \bmod p \end{aligned}$$

The inverse of d_1 exists since $d_1 x \equiv 1 \bmod p$ has a single solution for x by the fact that $d_1 < p$ is such that $(d_1, p) = 1$ and p is prime. $B = (d_1)^{-1} Z \bmod p$ is an integer in $0, \ldots, p-1$. a_1 is a (uniformly at) random (chosen) 0,1. What is the probability that the random a_1 is B? Naturally this probability is at most $1/2$, as after we fix B, a_1 can agree with this fixed value of B half of the time only. If the a_i's were not binary but ternary, then the probability bound would be $1/3$ instead. ■

Exercise 228.

Angluin-Valiant [2]. For every n, p, b with $0 \leq p \leq 1$ and $0 \leq b \leq 1$, we have the following.

$$\sum_{k=0}^{k=\lfloor (1-b)np \rfloor} B(n,p;k) \leq \exp(-b^2 np/2), \tag{4.54}$$

$$\sum_{k=\lceil (1+b)np \rceil}^{n} B(n,p;k) \leq \exp(-b^2 np/3), \tag{4.55}$$

where $B(n,p;k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Solution.
Omitted. See [2]. ■

Part II

Counting

Chapter 5

Combinatorics

5.1 Short review on counting principles

Counting involves counting configurations by developing formulas that describe the number of different configurations of objects.

5.1.1 Addition principle

If A and B are two disjoint sets of n and m objects, then their union C is of cardinality $n+m$.

Definition 5.1 **Addition principle**

If A is an event that can occur in n ways, and if B is an event that can occur in m ways, and the two events cannot occur at the same time, then event A or event B can occur in $n+m$ ways.

Call C the event that event A occurs or event B occurs.

5.1.2 Multiplication principle

If A and B are two (disjoint) sets of n and m objects, then their cartesian product $C = A \times B$ is of cardinality $n \cdot m$.

Definition 5.2 **Multiplication principle**

If A is an event that can occur in n ways, and if B is an event that can occur in m ways, then event A and event B can both occur in $n \cdot m$ ways.

Call C the event that A occurs and B occurs.

5.1.3 Generalizing the two principles

The multiplication and the addition principles can be generalized to more than two sets (events).

Example 5.1.1

Suppose that an event C occurs when events L, S, H all occur. Event L occurs by a student completing one of l language courses, event S occurs by a student completing one of s science courses, and event H occurs by a student completing one of h history courses, Event C is the satisfaction of a general education requirement that requires that a student completes one course from each one of the three areas (events). Find in how many ways event C can occur.

Solution. The answer is $l \cdot s \cdot h$. ■

Another way to formulate this example is as follows.

Example 5.1.2

Event D now is the satisfaction of a simple education requirement that requires that a student must take just one course out of the possible ones offered by the three areas with no restriction whatsoever whther the course is a language, science or history one. Find in how many ways even D can occur.

Solution. The answer is $l + s + h$. ■

5.1.4 Pigeonhole principle

If n pigeonholes are occupied by $n+1$ pigeons then at least one pigeonhole has more than one pigeon.

Definition 5.3

How many people do we need to gather in a room to make sure that at least two of them are born on the same day? Same day means the at least two are born on Mon, or Tue, or etc. What if same month? What if same date?

Solution. 8. 13. 32 ■

Definition 5.4

How many people do we need to gather in a room to make sure that at least four of them are born on the same day?

Solution. $3 \cdot 7 + 1$. If we have no four people born on the same day, in the worst case we have at most 3 born on every day, i.e. 21 people in the room to satisfy this. The 22nd will guarantee a four people born on the same day. ■

5.1.5 Permutations and k-permutations

Below the notion of a set of n objects implies that those n objects are distinct. "Given order" or "given sequence" implies that order matters: "1 2" is different from "2 1".

Definition 5.5 **Permutation**

Any arrangement of a set of n objects in a given order (sequence) is defined as a permutation of those objects.

We have an input pool of n objects. For an n-permutation we form n output positions (call them bins). The first (say left-most) position is filled by taking one of the n objects of the input. There are n choices. We place the object into the output's left-most position and remove it from the input pool. For the second (left-most) position of the output we then have $n-1$ choices among the $n-1$ objects left in the input pool and so on. The number of permutation of n objects is the number of n-permutations i.e. $P(n,n) = n!$.

Definition 5.6 **k-Permutation**

Any arrangement of any k of a set of n objects, $0 \leq k \leq n$, in a given order (sequence) is defined as a k-permutation of those objects.

We have an input pool of n objects. For an k-permutation we form k output positions (call them bins). The first (say left-most) position is filled by taking one of the n objects of the input. There are n choices. We place

the object into the output's left-most position and remove it from the input pool. For the second (left-most) position of the output we then have $n-1$ choices among the $n-1$ objects left in the input pool and so on until we reach the k-th left-most output position. The number of k-permutations of n objects is the number $P(n,k) = n \cdot (n-1) \cdot \ldots \cdot (n-(k-1))$, which is

$$P(n,k) = n \cdot (n-1) \cdot \ldots \cdot (n-(k-1)) = \frac{n!}{(n-k)!}.$$

Note that $P(n,k)$ is always an integer number.

Lemma 5.1 **Number of k-permutations and permutations**

The number of k-permutations of n distinct objects is given by the following expression.

$$P(n,k) = \frac{n!}{(n-k)!}.$$

The number of permutation is the number of n-permuations and is given by the following expression.

$$P(n,n) = P(n) = n!$$

Another way to generate the expression $P(n,k)$ or all the k-permutations is to move backwards from a permutation to a k-permutation. We will generate from a permutation of n objects, a k-permutation as follows. Given a permutation $x(1:n)$ with $x(i)$ being the object at position i from the left, isolate and move to the output and thus generate a k-permutation from $x(1:k)$. The issue is that several permutations have the common prefix $x(1:k)$ in the left-most k positions. How many exactly? The answer is $(n-k)!$ which is the number of permutations for $x(k+1:n)$ out of the $n-k$ objects not appearing in $x(1:k)$. Thus $(n-k)!$ n-permutations generate the same k-permutation. Because of this $P(n,k) = P(n)/(n-k)!$.

Observe that $P(n,k) = P(n-1,k) + kP(n-1,k-1)$.

5.1.6 Sample k out of n without replacement

Definition 5.7 **Sample k out of n without replacement**

The number of samples of k objects drawn from a pool of n objects without any replacement is equal to the following.

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = P(n,k).$$

5.1.7 Cyclic permutations

Definition 5.8 **Cyclic permutation**

Any arrangement of a set of n objects in a given circular (clock-wise) order is defined as a cyclic permutation of those objects.

A given order (sequence) has left-most and right-most position or start and finish. A cyclic permutation does not. If we lie it flat on the plane it can start from the first, second, etc position. That is it gives rise to n permutations.

Lemma 5.2 **Number of cyclic permutations**

The number of cyclic permutations of n distinct objects is given by the following expression.

$$\frac{P(n,n)}{n} = (n-1)!.$$

Thus $1,2,3$ and $2,3,1$ and $3,1,2$ map to the same circular permutation (call it $1,2,3$). Note that this gives a counter clock-wise sequence of permutations such as $1,3,2$ and $3,2,1$ and $2,1,3$ that are considered different and map clock-wise to $1,3,2$, a different circular permutation.

5.1.8 Permutations with repetitions

Definition 5.9 **Permutations with repetitions**

The number of permutations of n objects, with k_1 objects of one type, k_2 objects of another type, etc, k_l objects of another type is

$$P(n;k_1,k_2,\ldots k_l) = \frac{n!}{k_1!\cdot k_2!\cdot \ldots \cdot k_l!}$$

Example 5.1.3

Find all five letter words formed by the five letters of the word DADDY. (Improvise with another similar word!)

Solution. We are certain that the number of five letter words is NOT going to be $5! = 120$ but less. In fact they should be

$$\frac{5!}{3!\cdot 1!\cdot 1!} = 20.$$

The first 10 of them are shown below. The remaining ten of 10 are by swapping A and Y in them!

```
A Y ->  DDDAY , DDADY, DDAYD,
                DADDY, DAYDD, DADYD
                ADDDY, ADDYD, ADYDD, AYDDD
Y A ->  ? 10 more.
```

■

5.1.9 Derangements

Definition 5.10 **Derangement**

A derangement is a permutation of n objects such that no object occupies its initial position that is, it has no fixed points.

For example for $n = 3$ and numbers $\{1,2,3\}$ the permutation σ, $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$ is a derangement, while the permutation τ, $\tau(1) = 2, \tau(2) = 1, \tau(3) = 3$ is not a derangement, since 3 is a fixed point.
Thus if object o_i initially occupies position i, derangement (permutation) p maps o_i to $p(o_i)$ such that $p(o_i) \neq i$ for all i.

5.1.10 Counting subsets: k-of-n combinations

Note that in a subset, the order of the objects does not matter. Thus $\{1,2,3\} = \{3,2,1\}$.

Definition 5.11 **k-combination**

A k-combination of a set of n objects, where $0 \leq k \leq n$, is a subset containing k of those objects.

A k-combination is a subset of cardinality k (of n objects). Note that there is one n-combination of n objects, yet there exist $n!$ permutations of n objects.

Lemma 5.3 **Number of k-combinations of n objects**

The number of k-combinations of aset of n objects that is the number of subsets of cardinality k of a set of n objects is denoted by $C(n,k)$ and given by the following expression.

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Moreover,

$$C(n,k) = C(n-1,k) + C(n-1,k-1).$$

Note that for a set of n objects we assume that the n objects are distinct and so are the k objects that form a subset of those n objects!

We also say $C(n,k)$ is the number of combination of k out of n.

Note that several 3-permutations such as $1,2,5$, and $1,5,2$ and $2,1,5$ and $2,5,1$ an $5,1,2$ and $5,2,1$ generate one and the same 3-set $\{1,2,5\}$. There are $3!$ of them as many as the permutations of the three objects $1,2,5$.

We show how a k-permutation leads to a k-combination, and then we count the k-combinations by counting correctly and appropriately the k-permutation(s) leading to it.

To understand the derivation for $C(n,k)$, first obtain a k-permutation. $k!$ different k-permutations lead to the same k-object ubset. Thus to count the k-object subsets uniquely we need to divide the number of k-permutations $P(n,k)$ by $k!$. There are $P(n,k)$ k-permutations of n objects. $k!$ of them contain the same objects that lead to the same k-subset. In order to count the k-subsets correctly (only once), divide $P(n,k)$ by $k!$ to generate $C(n,k)$.

Alternatively, start with n objects, and obtain a k-permutation. In the k-permutation the order of the objects matters. In a combination (a set) the order does not matter. Thus $k!$ different k-permutations map to the same set, a k-combination. Thus $C(n,k) = P(n,k)/k!$.

Note that

$$C(n,k) = \binom{n}{k},$$

and it reads "n choose k" or "choose k out of n".

The second expression of the lemma can follow by combinatorial interpretations. Let the n objects be $e_1, e_2, \ldots, e_n$. We count the k-combinations of those n objects as follows by distinguishing two cases and using induction. Let the k-set to be formed is denoted by K.

Case 1: $e_1 \in K$. If $e_1 \in K$ it means K has $k-1$ other objects . Therefore $B = K - \{e_1\}$ is a $k-1$-combination and let us call it B. By induction the number of $(k-1)$-combinations that can form B out of the remaining $n-1$ objects (the original n minus e_1) is $C(n-1,k-1)$.

Case 1: $e_1 \notin K$. If $e_1 \notin K$ it means K has k objects out of $e_2, \ldots, e_n$. By induction the number of k-combinations that can form K out of the remaining $n-1$-objects (since e_1 is not in K) is $C(n,k-1)$.

The two cases leads to different k-combinations (i.e. different k-subsets) as one contains e_1 and the other does not. Therefore

$$C(n,k) = C(n-1,k-1) + C(n,k-1).$$

Use $C(i,j) = i!/(j!(i-j)!)$ to derive the desired bound.

Example 5.1.4

For a set $\{A,B,C\}$ Find and count all the 2-subsets (2-combinations).

Solution. We first find and count the two subsets that contain A. If A is one of the two objects of a two-subset the other object can be B or C. Thus there are two 2-subsets containing A. These are

$$\{A,B\} \qquad \{A,C\}.$$

If A is NOT oone of the two objects of a two-subset, then B and C must be to get a two-subset. Thus there is one 2-subsets not containing A. This is

$$\{B,C\}.$$

These are all: $C(3,2)=3$, as needed. ■

Example 5.1.5

For a set $\{A,B,C,D\}$ Find and count all the 2-subsets (2-combinations).

Solution. We generate first all 2-permutations and then filter them to generate distinct 2-combinations.

```
A B, B A,   A C, C A,   A D, D A,  B C, C B,  B D, D B,  C D, D C
   :           :           :          :          :          :
 {A, B}      {A, C}      {A, D}     {B, C}     {B, D}     {C,D}
```

$$C(4,2)=\frac{P(4,2)}{2!}=\frac{12}{2}=6.$$

■

Example 5.1.6

Suppose that an event C occurs when events L, S, H all occur. Event L occurs by a student completing two of 4 language courses, event S occurs by a student completing three of 5 science courses, and event H occurs by a student completing two of 6 history courses, Event C is the satisfaction of a general education requirement that requires that a student completes the sequence as specified. Find in how many ways event C can occur.

Solution. There are $\binom{4}{2}$ ways to satisfy the L requirement. There are $\binom{5}{3}$ ways to satisfy the S requirement. There are $\binom{6}{2}$ ways to satisfy the H requirement. The answer is as follows.

$$\binom{4}{2}\cdot\binom{5}{3}\cdot\binom{6}{2}.$$

■

5.1.11 Inclusion-Exclusion principle: two sets

We have two finite sets A and B with cardinalities $c(A)$ and $c(B)$. Their intersection has cardinality $c(A\cap B)$. The cardinality of the union is as follows.

$$c(A\cup B)=c(A)+c(B)-c(A\cap B)$$

Example 5.1.7

In a group of n students 20 speak French, 30 speak German and 5 of them speak both French and German. (All of them speak English.) Find n.

Solution. Every student who speaks French becomes member of set A. Every student who speaks German becomes a member of set B. Some student (five of them) are in both sets because they speak both languages. Consider A the first group $c(A)=20$, consider B the second group $c(B)=30$, and we also know that $c(A\cap B)=5$. Then

$$c(A\cup B)=c(A)+c(B)-c(A\cap B)=20+30-5=45.$$

■

5.1.12 Inclusion-Exclusion principle: three sets

We have three finite sets A, B, and C with cardinalities $c(A)$, $c(B)$, and $c(C)$.

$$c(A\cup B\cup C)=c(A)+c(B)+c(C)-c(A\cap B)-c(B\cap C)-c(C\cap A)+c(A\cap B\cap C).$$

Example 5.1.8

In a group of n students 20 speak French, 30 speak German and 5 of them speak both French and German. (All of these speak English.) There 55 students who speak only English (and neither French nor German). Find n.

Solution. Every student who speaks French becomes a member of set A. Every student who speaks German becomes a member of set B. Every student who speaks English becomes a member of set C. Consider A the first group $c(A)=20$, consider B the second group $c(B)=30$, consider C the third group $c(C)=100$, and we also know that $c(A\cap B)=5$, and we also know that $c(A\cap C)=20$, and we also know that $c(B\cap C)=30$. Implied is $c(A\cap B\cap C)=5$. Then

$$\begin{aligned} c(A\cup B\cup C) &= c(A)+c(B)+c(C)-c(A\cap B)-c(B\cap C)-c(C\cap A)+c(A\cap B\cap C)\\ &= 20+30+100-5-30-20+5=100. \end{aligned}$$

This problems requires the solution of the previous one first. ■

A member $a\in A$ is counted in $c(A)$. If it is also a member of B it has been counted in $c(B)$ as well, and thus we adjust by $-c(A\cap B)$ so that the member is not double counted. Likewise if a is part of C as well we adjust by $-c(C\cap A)$ but also $-c(B\cap C)$ if a is a member of B and C but not in A. However if a is part of all three sets we counted a three times in the three sets, and we subtracted its presence three times with $-c(A\cap B)-c(B\cap C)-c(C\cap A)$ so we adjust by adding $c(A\cap B\cap C)$, to count it at least (in fact, exactly) once.

Example 5.1.9

In a group of n students 20 speak French, 30 speak German, and 100 English. 20 speak French and English, 30 speak German and English, and 5 speak French and German. There are 4 who speak all three languages. Find n.

Solution. Every student who speaks French becomes a member of set A. Every student who speaks German becomes a member of set B. Every student who speaks English becomes a member of set C. Consider A the first group $c(A)=20$, consider B the second group $c(B)=30$, consider C the third group $c(C)=100$, and we also know that $c(A\cap B)=5$, and we also know that $c(A\cap C)=20$, and we also know that $c(B\cap C)=30$. Explicitly it is given $c(A\cap B\cap C)=4$. Then

$$\begin{aligned} c(A\cup B\cup C) &= c(A)+c(B)+c(C)-c(A\cap B)-c(B\cap C)-c(C\cap A)+c(A\cap B\cap C)\\ &= 20+30+100-5-30-20+4=100. \end{aligned}$$

■

5.1.13 Inclusion-Exclusion principle: more sets

Let n sets be $A_1,\ldots,A_n$ and let us define,

$$\begin{aligned} k_1 &= \sum_i c(A_i),\\ k_2 &= \sum_{i<j} c(A_i\cap A_j),\\ k_3 &= \sum_{i<j<k} c(A_i\cap A_j\cap A_k),\\ &\cdots \end{aligned}$$

Then

$$c(A_1 \cup A_2 \cup \ldots \cup A_n) = k_1 - k_2 + k_3 - \ldots + (-1)^{m-1} k_m$$

Moreover,

$$c(A_1^{\complement} \cap A_2^{\complement} \cap \ldots \cap A_n^{\complement}) = c((A_1 \cup A_2 \cup \ldots \cup A_n)^{\complement}) = |U| - k_1 + k_2 - k_3 + \ldots + (-1)^m k_m$$

A proof of the first formula can follow if one considers object e. If object e is in t of those sets, then in k_1 this object e is counted t times or in other words $C(t,1)$ times ($C(n,k)$ is the number of k combinations of n objects). Moreover in k_2 it is it is counted $C(t,2)$, and so on. Therefore in the right hand side of $c(A_1 \cup A_2 \cup \ldots \cup A_n)$ the count of e is

$$C(t,1) - C(t,2) + C(t,3) - \ldots + (-1)^{t-1} C(t,t).$$

But from the binomial theorem

$$0 = (1-1)^t = \sum_{i=0}^{t} C(t,i)(-1)^i (1)^{t-i} = C(t,0) - C(t,1) + C(t,2) - C(t,3) + \ldots.$$

Solving for $C(t,0)$ we have the following.

$$1 = C(t,0) = C(t,1) - C(t,2) + C(t,3) - \ldots + (-1)^{t-1} C(t,t).$$

Therefore the total count for e in the sum of the Inclusion-Exclusion's formula is 1!

5.1.14 Partitions

Definition 5.12 **Partitions**

For a set of n distinct object we partition them into m labeled disjoint partitions (sets). This way each set contains $k_1, k_2, \ldots, k_m$ of the objects respectively where $k_1 + \ldots + k_m = n$.

Whether the first partition with say $k_1 = 2$ is $\{1,2\}$ or $\{2,1\}$ does not matter as both refer to the same partition (subset).

Lemma 5.4 **Number of ways to partition n objects into m partitions**

The number of way to partition n objects into m partitions so that the i-th partition gets k_i objects, $1 \leq i \leq m$, and $k_1 + \ldots + k_m = n$ is given below.

$$P(n; k_1, \ldots, k_m) = \binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1! k_2! \ldots k_m!}.$$

Proof. For the first partition there are $\binom{n}{k_1}$ ways to pick its k_1 objects. For the second partition there are $\binom{n-k_1}{k_2}$ ways to pick its k_2 objects, and so on for the m-th partition there are $\binom{n-k_1-\ldots-k_{m-1}}{k_m}$ ways to pick its k_m objects with the total number of partitions described being equal to

$$\begin{aligned}
\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\ldots-k_{m-1}}{k_m} &= \frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdots\frac{(n-k_1-\ldots-k_{m-1})!}{k_m!(n-k_1-\ldots-k_{m-1})!} \\
&= \frac{n!}{k_1! k_2! \ldots k_m!} \\
&= \binom{n}{k_1, k_2, \ldots, k_m}.
\end{aligned}$$

■

A k-combination is a partition with $m = 2$ and $k_1 = k$ and $k_2 = n - k$.

Example 5.1.10

There are six toys A through F that need to be split to three kids, so that each kid gets two of them. Find the number of ways we can split them.

Solution.

$$\binom{6}{2,2,2} = \frac{6!}{2!\cdot 2!\cdot 2!} = 6!/8 = 90.$$

■

Example 5.1.11

Find the coefficient of $x_1^{k_1}\dots x_m^{k_m}$ in the expansion of

$$(x_1+x_2+\dots+x_m)^n,$$

where $k_1+\dots k_m = n$.

Solution.

$$\binom{n}{k_1,k_2,\dots,k_m} = \frac{n!}{k_1!k_2!\dots k_m!}.$$

■

5.1.15 Selection and Multisets : combinations with repetitions

Definition 5.13 **k-multiset, k-selection**

A k-multiset or k-selection is an arrangement of k objects drawn with repetition from a collection of n available types of objects.

Say we have three types of marbles: red, green and blue, and we would like to draw five marbles out of them. Then $n=3$ and $k=5$.

Suppose we have an urn with n objects labeled 1 through n. We draw an object and account of its label. The object is then repositioned into the urn. We draw the next object from the urn and account of its label. After k drawings, we account how many times the object with label 1 was drawn, the object with label 2, and so on.

Lemma 5.5 **Number of k-multisets of n types**

The number of ways $S(n,k)$ we can select k objects drawn with repetition from a collection of n available types of objects is as follows.

$$S(n,k) = \binom{n+k-1}{k}.$$

Proof. On a line we place $n+k-1$ markers $-$. Of those markers, k of them will be turned into o, and the remaining $n-1$ into a $|$. The o would represent an object among the k selected. If an o is on the left of the left-most $(i=1)$ $|$ it is of type 1. If an o is on the right of the right-most $(i=n-1)$ $|$ it is of type n. If an o is on the right of the i $|$ and the $i+1$ $|$ it is of type $i+1$. The number of ways to turn k of the $n+k-1$ dashes $-$ into o is equal to

$$C(n+k-1,k) = \binom{n+k-1}{k},$$

the remaining $n-1$ dashes $-$ become $|$ and thus determine the type of an o. ■

We can read $\binom{n+k-1}{k}$ as "n multichoose k".

Example 5.1.12

We have 5 types of an object. We draw 6 objects. In how many different ways can we draw those 6 objects?

Solution.

$$C(n+k-1,k) = C(5+6-1,6) = C(10,6).$$

Two possibilities (configurations) are shown:

```
o o | o | | o o | o    2xType1, 1xType2, 2xType4 , 1xType5
o | o o | o | o | o    1xType1, 2xType2, 1xType3,  1xType4, 1xType5
```

■

5.1.16 Integer equation: number of non-negative solutions

Lemma 5.6 **Sum k integer equation with non-negative solutions**

The integer equation with $x_i \in \mathbb{N}$

$$\sum_{i=1}^{n} x_i = x_1 + \ldots + x_n = k \quad \text{with} \quad x_i \geq 0$$

has $C(n+k-1,k)$ non-negative integer solutions.

Solution. Omitted. ■

5.1.17 Integer equation: number of positive solutions

Lemma 5.7 **Sum k integer equation with positive solutions**

The integer equation with $x_i \in \mathbb{N}$

$$\sum_{i=1}^{n} y_i = y_1 + \ldots + y_n = k \quad \text{with} \quad y_i > 0$$

has $C(n+k-1,k)$ non-negative integer solutions.

Consider now another problem. Find the number of solutions to the equation

$$\sum_{i=1}^{n} y_i = y_1 + \ldots + y_n = k \quad \text{with} \quad y_i > 0,$$

has $C(k-1,k-n)$ positive integer solutions.

Solution. Since $y_i > 0$ then set $y_i = x_i + 1$, with $x_i \geq 0$. Note that if $y_i > 0$ then $k \geq n$ as well. We can reduce this question to the previous integer equation question. ■

5.1.18 Choosing k objects from n with repetitions: order matters

Definition 5.14 **Choose k from n with repetitions**

The number of ways of choosing k objects out of n distinct objects, where repetitions are allowed is n^k. (Order matters.)

Lemma 5.8 **Choose k from n with repetitions**

The number of ways of choosing k objects out of n distinct objects, where repetitions are allowed is n^k. (Order matters.)

The first of the k objects is one of the n available. After we draw say object i, a copy of object i replenishes the pool of $n-1$ objects (where i is missing) to n. This repeats a total of k times.

Example 5.1.13

The number of 3 digit integers, where a digit takes 10 different values is 10^3.

Example 5.1.14

The number of m digit binary numbers, where a digit takes 2 different values (0 and 1) is 2^m.

If order does not matter the answer is $S(n,k)$.
Equivalently, we have the following interpretation.

5.1.19 Sample k out of n with replacement: order matters

Lemma 5.9 **Sample k from n with replacement**

The number of ways of choosing k samples with replacement out of a pool of n distinct objects, is equal to

$$n \cdot n \cdot \ldots \cdot n = n^k.$$

In both cases order matters. If order does not matter the answer is $S(n,k)$.

Example 5.1.15

(a) We have a dice (six faces 1 through 6) and is thrown three times. How many different outcomes?
(b) What are the chances of getting 5 three times in a row?

Solution.
(a) $6 \cdot 6 \cdot 6$ including a 555.
(b) $\frac{1}{6^3}$. ■

5.1.20 Choosing k out of n

We summarize below four of the cases mentioned earlier.

Choose k of n	Ordered (sequence)	Unordered (set)
No repetitions	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$
With repetitions	n^k	$\binom{n+k-1}{k}$

5.1.21 Partition identical objects into k possibly empty groups

Definition 5.15 **Partitioning n identical objects into k groups**

We have n identical objects. We would like to assign distinct k labels to those objects. Think of placing those objects into k boxes, each box of a different label.

Lemma 5.10 **Number of partitions of n id. objects into k groups**

The number of ways of partitioning n identical objects into k groups is equal to the following expression

$$C(n+k-1,n) = C(n+k-1,k-1).$$

Proof. Similarly to a prior problem we place down $n+k-1$ markers marked $-$. Next we place $k-1$ markers marked $|$; these generate k groups (call them boxes). The remaining n markers are turned into o. The number of o on the left of the first marker, or between two markers, or to the right of the right-most marker indicate the number of identical objects assigned to each one of the k groups. The number of partitions is then

$$C(n+k-1,k-1) = C(n+k-1,n).$$

■

5.1.22 Partition identical objects into k non-empty groups

Definition 5.16 **Partitioning n identical objects into k non-empty groups**

We have n identical objects. We would like to assign distinct k labels to those objects, each k label assigned at least once. Think of placing those objects into k boxes, each box of a different label.

Lemma 5.11 **Number of partitions of n id. objects into k groups**

The number of ways of partitioning n identical objects into k non-empty groups is equal to the following expression

$$C(n-1,k-1) = C(n-1,n-k).$$

Proof. We know every group will get at least one object. Thus $n \geq k$. Instead of starting with $n+k-1$ markers, we start with k fewer that is $n-1$ markers marked $-$. $k-1$ of them will be turned $|$, and the remainder o. When counting the number of o's for a group add to the total a $+1$. ■

5.1.23 Partition identical objects into k non-empty indistinguishable groups

Definition 5.17 **Partitioning n identical objects into k identical non-empty groups**

We have n identical objects that are to be split into k indistinguishable groups. This is the problem of partitioning integer n into k groups (parts).

Lemma 5.12 **Number of Partitions of n identical objects into k identical non-empty groups**

The number of partitions $p_k(n)$ of n objects into k groups is equivalent to the placement of $n-k$ objects into those k groups since the k groups are not to be empty and is given by the following expression.

$$p_k(n) = p_{k-1}(n-1) + p_k(n-k),$$

where $p_1(n) = 1$, $p_n(n) = 1$, and $p_k(n) = 0$ for $k > n$.

Proof. We move backwards to count and calculate $p_k(n)$. If one group has exactly one object, we remove that object, we are done with that group, we are left with $n-1$ objects and $k-1$ other groups thus the number of ways of partitioning the remaining $n-1$ objects into those $k-1$ remaining groups is, by induction, $p_{k-1}(n-1)$. Otherwise all groups have at least 2 objects, we subtract one object from each group, we still have k groups with at least one object, the number of objects left is $n-k$ and thus the number of partitions of those $n-k$ left-over objects over the k groups is $p_k(n-k)$. Adding the two contributions, the result follows. ■

5.2 Addition and Multiplication principle: examples

Exercise 229.

Addition principle example. A student needs to choose a Science elective i.e. either a physics or a biology course. The Physics Department has 3 appropriate ones, the Biology Department has 4. How many choices for the student?

Solution.
The student has 7 choices.

■

Exercise 230.

Addition principle example. Suppose that YWCC has 10 CS courses, 5 IS course, and 3 IT courses. What is the number of ways a student can pick a CS/IS/IT elective?

Solution.
The number of ways a student can pick a CS/IS/IT elective course (CS or IS or IT) is $10+5+3=18$. ■

Exercise 231.

Muliplication principle example. Suppose that YWCC has 10 CS courses, 5 IS course, and 3 IT courses. What is the number of ways a student can choose just one course per program for a total of three courses? (There are three programs indicated: CS, IS, and IT.)

Solution.
The number of ways a student can choose one course from each program (for a total of three courses) is $10 \cdot 5 \cdot 3 = 150$. ■

Exercise 232.

Addition and Multiplication Principle example. A digit is one of 0-9. A lower-case char is one of a-z. An upper-case char is one of A-Z. A char is upper-case or lower-case. An alphanumeric digit is a char or a digit.
(a) How many choices for a char?
(b) How many choices for an alphanumeric digit?
(c) How many choices for a two digit number?

Solution.
(a) A char is upper-case or lower-case. By the addition principle we have $26+26=52$ possible outcomes for a char.
(b) An alphanumeric digit is a char or a digit. By the addition principle we have $26+26+10=62$ possible outcomes.
(c) A two digit number by the multiplication principle has 100 outcomes. (Leading zeroes considered.) ■

Exercise 233.
A class has 20 students 12 CS and 8 IT.
(a) A class rep. is to be elected. How many ways to elect one?
(b) 2 class rep. are to be elected, one CS and one IT. How many ways to elect one?
(b) 2 class rep. are to be elected, one President one Vice-P. How many ways to elect such a pair?

Solution.
(a) By the addition principle $12+8=20$.
(b) By the multiplication principle $12\cdot 8=96$.
(c) We 20 choice for the President and 19 for the Vice-P for a total of $20\cdot 19=380$. ■

Exercise 234.
Given 3 English books, 5 CS book, and 7 Math books, in how many ways can we choose 2 books in different topics?

Solution.
$3 \cdot 5 + 5 \cdot 7 + 3 \cdot 7$. ■

Exercise 235.
We throw two dice (the plural of a die is dice). What are the chances of getting an even number? A die has six faces each one mapping to 1 through 6; the corresponding value is indicated by a number of dots on the correspond face. The plural for die is dice.

Solution.
Out of the 36 possible draws (6 per die, 2 dice thrown) half are odd-odd or even-even leading to an even sum. That is 18 out of 36 i.e. 50%. ■

Exercise 236.
How many different bytes do we have? A byte is a sequence of 8 bit. A bit is 0 or 1.

Solution.
By the addition principle we have 2 possibilities for a bit. By the multiplication principle we have $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 256$ possibilies. Thus a byte has 256 possible values and thus we have 256 different bytes. ■

Exercise 237.
Show that a set with n elements has 2^n subsets.

Solution.
Consider n-bit $b = b_1 b_2 \dots b_n$ sequence with $b_i = 1$ if element i is in the set and $b_i = 0$ otherwise. Extend the previous problem to obtain 2^n choices for a b (extend from eight to n). Each b of the 2^n possible ones indicates a different subset. ■

Exercise 238.

Deck of cards. A deck of cards (standard deck) has 52 cards split into four suits of clubs, diamonds, hearts and spades. Each suit has 13 cards with ranks from ace (one), two, etc to ten, jack (J), queen (Q), and king (K). Hearts and Diamonds are colored red, Clubs and Spades are colored black.
How many ways can we deal a King or black card from a standard deck of cards if you are served one card?

Solution.
There are 26 black card (clubs and spades) plus two Kings among the read cards (hearts and diamonds). This is 28 total out of 52. ■

Exercise 239.
A student needs to choose a Science elective as before and a course from the Math Department. The Physics Department has 3 appropriate ones, the Biology Department has 4, and the Chemistry Department 2. The student has 9 choices. There are five choices for a Math course, and also two choices for a Statistics course from the Math Department. How many possibilities does the student have?

Solution.
Now the combination of 1 Science Dept and 1 Math Dept course can occur in $(3+4+2)\times(5+2)=63$ ways. ■

Exercise 240.
We have n objects say $\{1,2,\ldots,n\}$. What is the number of subsets of this set of n objects?

Solution.
Already solved! Let s_i map to choosing or not choosing i. We have 2 choices for s_i. For all s_i i.e. for the number of subsets by the multiplication principle we have $2 \cdot 2 \cdot \ldots \cdot 2 = 2^n$. ■

Exercise 241.
How many five digit (non-negative) integers are there with no leading zero, that are even?

Solution.
The number of five digit integers with no leading zero are

$$9 \cdot 10^4.$$

Half of them are even, half of them are odd. Complete the details. ■

Exercise 242.
(a) Five people are seated in a row of chairs. What is the number of ways that they can be seated?
(b) Five people are seated on the chairs of a round table. What is the number of ways that they can be seated?

Solution.
5! and 4! respectively. ■

Exercise 243.
(a) Find the number of distinct words (it does not matter if they do not map to an English word) that can be formed by the letters of the word MOTHER.
(b) Find the number of distinct words (it does not matter if they do not map to an English word) that can be formed by the letters of the word MOMMY.

Solution.
(a) $6!$.
(b) $\frac{5!}{3!} = 20$. (See the DADDY example.) ■

Exercise 244.
A class has 12 CS and 8 IT students.
(a) In how many way can we form a team of 4 students?
(b) In how many way can we form a team of 4 students, 2 of which are CS and 2 IT?
(c) Elect a President, VP, Secretary, and Treasurer?

Solution.
(a) $\binom{20}{4}$.
(b)

$$\binom{12}{2} \cdot \binom{8}{2}.$$

(c) $P(20,4)$. ■

Exercise 245.
A class has 12 CS and 8 IT students.
(a) In how many way can we form a team of 4 students, where one of the four is the designated team leader?

Solution.
(a) $20 \cdot \binom{19}{3}$. ■

Exercise 246.
A "password" has five alpharithmetic characters, the first two are alphabetic (A-Z, a-z only) and the last three are numeric (0-9 each).
(a) How many passwords in total to crack?
(b) How many passwords with distinct alphabetic chars?
(c) How many passwords with distinct alphabetic chars and distinct numeric chars?
(d) How many passwords with distinct alphabetic chars and distinct numeric chars but no 0 as the first digit (012 not allowed but 203 or 230 is ok)?

Solution.
(a) $26 \cdot 26 \cdot 10^3$.
(b) $26 \cdot 25 \cdot 10^3$.
(c) $26 \cdot 25 \cdot 10 \cdot 9 \cdot 8$.
(d) $26 \cdot 25 \cdot 9 \cdot 9 \cdot 8$. ■

Exercise 247.
(a) In how many ways can 5 people sit in a row (of 5 chairs)?
(b) In how many ways can 5 people sit in a row (of 5 chairs), under the condition 1 and 2 are seated next to each other.
(c) Repeat a and b for a circular table.

Solution.
(a) $5! = 120$.
(b) 1 and 2 can be seated next to each other as $1,2$ in 4 different ways and $2,1$ in 4 more, a total of 8. The remaining 3 in $3!$ ways. The total is $4 \cdot 2 \cdot 3! = 48$.
(c) $4!$.
(d) Think clock-wise $1,2$ are next to each other. Then 3,4,5 are arranged after 2 in $3!$ ways. Or $2,1$ similarly. Total 12. ■

Exercise 248.
In how many ways can we flip a coin $2n$ times so that now two heads appear in a row and we have n heads (and of course n tails)?

Solution.
We must have a patter $HTHT\dots H$ with n heads and $n-1$ tails for a total of $2n-1$. The n-th tail can fall before the first H, etc, before the n-th H, or after the n-th H, a total of $n+1$ possibilities. ■

Exercise 249.
How many five digit (non-negative) integers are there with no leading zero, that are even, and no digit repeats itself?

Solution.
Let $a_5a_4a_3a_2a_1$ be the number. We have 9 choices for a_5 as before. We start from the right before we go left to right. There are 5 choices for a_1. Then we have 8 choices for a_5. We have 8 choices for a_4 (10 minus the a_5 and a_1 choices). Then 7 for a_3, 6 for a_2, for a total of

$$5 \cdot 8 \cdot 8 \cdot 7 \cdot 6.$$

■

5.3 Pigeonhole examples

Exercise 250.
(a) In a class of 8 students, determine that two students were born on the same day of the Week.
(b) In a class of 13 students, determine that two students were born on the same month.
(c) In a class of 32 students, determine that two students were born on the same date. (We are talking about Eartch months not Martian months that are longer than 31 days.)
(d) In a group of 101 people, determine that two people were born on the same date. (It assumes that it is rare to encounter a centenarian.)
(e) In a group of 367 people, determine that two people share the same birthday. (It assumes that it is rare to encounter a Martian.)

Solution.
(a) Straightforward by the principle. (We don't know what day that might be though unless we ask.)
(b) Straightforward by the principle. (We don't know what month that might be though unless we ask.)
(c) Straightforward by the principle. (We don't know what date that might be though unless we ask. Furthermore, we are talking about Earth months not Martian months that are longer than 31 days.)
(d) Easy.
(e) Easy. ■

5.4 Inclusion-exclusion: examples

Exercise 251.
In a class of CS610, 10 students know C, 20 students know C++, and 30 students know Java. 15 students know C++ and Java, 5 students know C and Java and 8 students know C and C++. 3 students know all three. What is the total number of students in the class?

Solution.
Let A be the C students, B be the C++, and C the Java students.

$$\begin{aligned} c(A \cup B \cup C) &= c(A)+c(B)+C(C)-c(A \cap B)-c(B \cap C)-c(C \cap A)+c(A \cap B \cap C) \\ &= 10+20+30-15-5-8+c(A \cap B \cap C) \\ &= 60-28+3=35 \end{aligned}$$

■

Exercise 252.
In a class of CS610 with 50 students, 30 students know C and 36 students know Java. Show that are at least 15 students who know C and Java.

Solution.
Let X denote the students who know C. Then $|X| = 30$. Let Y denote the students who know Java. Then $|Y| = 36$. Then $X \cap Y$ be the students who know C and Java. We are going to show $|X \cap Y| \geq 20$. We have by inclusion exclusion

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

This implies the following.

$$|X \cup Y| = 30 + 36 - |X \cap Y|.$$

But $|X \cup Y| \leq 50$.

$$50 \geq |X \cup Y| = 30 + 36 - |X \cap Y|.$$

Therefore

$$|X \cap Y| \geq 66 - 50 = 16 \geq 15.$$

■

Exercise 253.
Find the number of natural numbers (positive integers greater than 0) less than or equal to 100 which are divisible by 2 or 5.

Solution.
50 of them are divisible by 2 (even numbers). 20 of them are divisible by 5. But there are 10 numbers among the 50 of the first group which are part of the second group and thus double counted in the sum $20+50=70$. These are the multiples of 10, a total of 10!. Thus the number of natural numbers divisible by 2 or 5 is $50+20-10=60$. In other words add to 50 the multiples of 5 which are not even $(5,15,25,\ldots,95)$! ■

Exercise 254.
How many of the integers 1,2,...,100, are divisible by at least one of 2, 3, or 7? Don't use brute force to answer this problem. Justify your answer.

Solution.
Apply the inclusion/exclusion principle. If S_n is the number of integers between 1 and 100 divisible by n, the answer is

$$S_2 + S_3 + S_7 - S_6 - S_{14} - S_{21} + S_{42} \quad = \quad 50 + 33 + 14 - 16 - 7 - 4 + 2 \quad = \quad 72$$

■

Exercise 255.
Find the number of prime numbers between 2 and 100 inclusive. There are 99 integers between 2 and 100 inclusive.

Solution.
An $2 \leq n \leq 100$ is prime if it has no prime factor ≤ 10 (or it is one of those prime numbers). The only prime numbers ≤ 10 are 2,3,5, and 7. Then, we have, where $A_i^{\complement}$ is the set of integers not a multiple of i,

$$\begin{aligned} c(A_2^{\complement} \cap A_3^{\complement} \cap A_5^{\complement} \cap A_3^{\complement}) &= 99 - (100/2) - (100/3) - (100/5) - (100/7) + (100/2 \cdot 3) + (100/2 \cdot 5) + (100/2 \cdot 7) \\ &+ \quad + (100/3 \cdot 5) + (100/3 \cdot 7) + (100/5 \cdot 7) - (100/2.3.5 + 100/2.3.7 + 100/2.5.7) + 100/3.5.7) + (100/(2.3.5.7 \\ &= 99 - 50 - 33 - 20 - 14 + 16 + 10 + 7 + 6 + 4 + 2 + 3 + 2 + 1 + 0 - 0 = -18 + 45 - 5 = 21. \end{aligned}$$

Adding to the sum the 4 prime numbers 2,3,5, and 7 we get a total of 25. ■

Exercise 256.
Let U be the set of positive integers up to and including 100. Find the number of them which are not divisible by 3,5,7.

Solution.
The set A containing integers divisible by 3 hass $c(A) = 33$. The set B containing integers divisible by 5 has $c(B) = 20$ and likewise for the sect C of integers divisible by 7, $c(C) = 14$. Also $c(A \cap B) = 6, c(A \cap C) = 4, c(B \cap C) = 2$. Moreover $c(A \cap B \cap C) = 0$. Thus

$$c(A^{\complement} \cap B^{\complement} \cap C^{\complement}) = 100 - (33 + 20 + 14) + (6 + 4 + 2) - 0 = 100 - 67 + 12 = 45.$$

■

Exercise 257.
Let A be a set of objects $1,\ldots,n$. Let P_i be properties on the objects of A and let the number of such properties be m. Let $A(i,j,\ldots,k)$ be the number of elements of A which have properties $P_i,P_j,\ldots,P_k)$. (This include $A(i)$, $A(i,j)$, $A(i,j,\ldots)$ etc.) Then the number of elements of A having at least one of these properties is given by the following expression

$$N=\sum_i A(i)+(-1)^1\sum_{i<j}A(i,j)+(-1)^2\sum_{i<j<k}A(i,j,k)+\ldots$$

Solution.
An element of A that has none of the listed properties contributes 0 to the sum N above. Consider an object u that has k of these m properties. Object u contributes m to the first sum $\sum_i A(i)$, $\binom{m}{2}$ to the second sum $\sum_{i<j}A(i,j)$, $\binom{m}{3}$ to the third sum and so on. Thus the total contribution of u is

$$m-\binom{m}{2}+\binom{m}{3}-\ldots+(-1)^{j-1}\binom{m}{j}=1-\left(1-m+\binom{m}{2}-\binom{m}{3}+\ldots-(-1)^{j-1}\binom{m}{j}+\ldots-(-1)^{m-1}\binom{m}{m}\right)=1-\sum_{j=0}^{m}\binom{m}{j}(-1)^j=$$

exactly one. The result follows. ■

Exercise 258.
Let A be a set of objects $1,\ldots,n$. Let P_i be properties on the objects of A and let the number of such properties be m. Let $A(i,j,\ldots,k)$ be the number of elements of A which have properties $P_i,P_j,\ldots,P_k)$. (This include $A(i)$, $A(i,j)$, $A(i,j,\ldots)$ etc.) Then the number of elements of A having LESS than one of these properties i.e. having exactly 0 of these properties is given by the following expression

$$|A|-N=|A|-\sum_i A(i)+(-1)^2\sum_{i<j}A(i,j)+(-1)^3\sum_{i<j<k}A(i,j,k)+\ldots$$

Solution.
Obvious. ■

5.5 Permutations: examples

Exercise 259.

Permutations. Any arrangement of a set of n objects in a given order is defined as a **permutation** of those objects. There are $P(n) = n!$ permutations of n objects. Prove the claim.

Solution.
Let the objects be $1, 2, \ldots, n$ and we want to fill $x_1 x_2 \ldots x_n$. There are n choices (possibilities) for x_1. Having committed on one, there are only $n-1$ possibilities for x_2 and so on. For x_1 there is only one possibility. The number of permutations possible is $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$. ■

Exercise 260.

Cyclic permutations.

How many cyclic permutations on n objects?

Solution.

$(n-1)!$. Consider the $n!$ permutations that can be arranged on a line (linear). In a cyclic permutation. There is no starting position. Thus a linear permutation $a_1a_2a_3a_4a_5\ldots a_n$ maps to the same cyclic permutation as $a_2a_3a_4a_5\ldots a_{n-1}a_na_1$, $a_3a_4a_5a_6\ldots a_na_1a_2$, $a_4a_5a_6a_7\ldots a_1a_2a_3$, $\ldots$, $a_na_1a_2a_3\ldots a_{n-3}a_{n-2}a_{n-1}$. Thus divide $n!$ by n possible starts discarded to get the answer. ■

Exercise 261.
How many permutations of the word ANNA are there?

Solution.
If we had ABCD the answer would be $4! = 24$. But $A_1N_1N_2A_2$ (subscript indicating the first or second instance) is the same as $A_2N_1N_2A_1$, and of course $A_1N_2N_1A_2$, $A_2N_2N_1A_1$. Thus $4!/(2! \cdot 2!) = 6$ is the answer. ■

Exercise 262.

k-**Permutations.** Let $P(n,k) = n(n-1)(n-2)\dots(n-k+1)$. Any arrangement of any k of a set of n objects, $k \leq n$ in a given order is defined as a **k-permutation** of those objects. There are $P(n,k) = n!/(n-k)!$ k-permutations of n objects. Prove the result. Also show the following

$$P(n,m) = P(n-1,m) + kP(n-1,m-1),$$

with $0 < m < n$.

Solution.
Think of filling k Rooms $R_1, R_2, \dots, R_k$ with one person each; each person is labeled $1, \dots, n$. For the first room R_1 we have n choices; when we move to R_2 we are left with $(n-1)$ choices since one of n has already booked into R_1. For R_3 we have only $n-2$ choices, and so on. Thus for room R_i we have $(n-i+1)$ choices. By the multiplication principle for the k rooms $R_1, R_2, \dots, R_k$ i.e. $\prod_{i=1}^{i=k}(n-i+1) = n(n-1)\dots(n-k+1)$.
We use m instead of k for the alternative formulation. Pick an arbitrary element t. t can be in the m-arrangement or not. If t is NOT in the arrangement the number of ways to create an m arrangement by induction is $P(n-1,m)$. If t is IN the arrangement out of the $n-1$ remaining keys we are going to pick (by induction) $P(n-1,m-1)$ arrangements. The $m-1$ objected picked create m slots; one of those m slots will be occupied by t. The two sets of permutations, not containing t and containing t are disjoint; use the addition principle to conclude the following.

$$P(n,m) = P(n-1,m) + mP(n-1,m-1).$$

■

Exercise 263.

Cyclic permutations. How many ways can we arrange n objects over a circle?

Solution.
$(n-1)!$. Each of $n!$ permutation can be circularly rotated n ways and maps to the same cyclic permutation. ■

Exercise 264.
(a) Find the number of 3-permutations of a, b, c.
(b) Find the number of 2-permutations of a, b, c.

Solution.
(a) It is $3! = 6$. The first element can be picked in three different ways (a, b, c) the second element in two and the third element in one ways after the first two elements have been picked. Total is $3 \cdot 2 \cdot 1 = 3! = 6 = P(3,3) = P(3)$.
(b) It is $3 \cdot 2 = 6 = P(3,2)$. ■

Exercise 265.
The number of permutations of n objects that are not distinct and there are n_1 instances of one type, n_2 of another type, ..., n_k of another type is $P(n;n_1,\ldots n_k)$

$$P(n;n_1,\ldots n_k) = \frac{n!}{n_1!n_2!\ldots n_k!}$$

Solution.
Treat the objects distinct first. There are $n!$ permutations p. But there n_1 objects of the first type and are thus not distinct. Any of the $n_1!$ ordering of those objects in p is the same as any other one. Thus we adjust by dividing $n!$ by $n_1!$ and we have reduce the number of permutations down to $n!/n_1!$. We continue similarly. ■

Exercise 266.
Let A,B,C be three objects. How many permutations of them do we have?

Solution.
We have 3! permuations of them namely, ABC, ACB, BAC, BCA, CAB, CBA. For three objects all permutations are of the pattern xyz. There are three choices for x (A,B, or C). Having picked a choice for x, there $3-1$ choices for y (the remaining two not used for x). Having picked a choice for x,y there is only one choice left for z. Thus $3\cdot 2\cdot 1 = 3!$. For the general case with n objects, induction works better. $xP(n-1)$. For the first letter with n choices. The remaining $n-1$ positions $P(n-1)$ will be filled by the permutation of the remaining $n-1$ objects which by induction is $(n-1)!$. Thus the total number of permutations of n objects is $n\cdot(n-1)! = n!$. ■

Exercise 267.

Let A, B, C be three objects. How many 2-permutations of them do we have? A 2-permutation means a sequence of two of those three objects.

Solution.

We have the following 2-permutations of those three objects: AB, BA, AC, CA, BC, CB Suppose we have n objects and we want to find the k permutations of them. Each k permutation is of the pattern $p_1 p_2 \dots p_k$. For p_1 there are n choices of the n objects. Having picked p_1, we are left with $n-1$ choices for p_2, and having picked a choice for p_2 we are left with $n-2$ choice for p_3 and working this out similarly we are left with $(n-k+1)$ choices for p_k. Thus the number of k-permutations of n objects is $P(n,k) = n \cdot (n-1) \cdot (n-k+1) = n!/(n-k)!$. Moreover $P(n,n) = n!$ so it works also for permutations (which are n-permutations)! ■

Exercise 268.
(a) Number of license plates with three letters and three digits.
(b) Same as in (a) but neither letter nor digits repeat.

Solution.
(a) Answer is $26 \times 26 \times 26 \times 10 \times 10 \times 10$.
(b) Answer is $26 \times 25 \times 24 \times 10 \times 9 \times 8$. ■

Exercise 269.
(a) We throw a die six times. How many outcomes?
(b) We throw a die six times. How many outcomes?

Solution.
(a) $6 \times 6 \times 6 \times 6 \times 6 \times 6$. ■

Exercise 270.
Permutations with repetitions. Consider you are given string YWCC. What are all possible four-letter words than can be formed?

Solution.

$$P(4;1,1,2) = \frac{4!}{1!1!2!} = 3 \cdot 4 = 12$$

YWCC, YCWC, YCCW, WYCC, WCYC, WCCY, CYWC, CWYC, CWCY CYCW, CCYW, CCWY. ■

Exercise 271.
How many circular permutations on A,B,C ?

Solution.
We fix A in position 1. Then we have two choices for B,C either BC or CB. Thus ABC and ACB are the circular permutations. In other wordds, For the three letter A,B,C set let us consider the 3! permutation answer. Three of those permutations: ABC, CAB and BCA are the same circular permutation. So are the other three ACB, BAC, and CBA. Thus There are only two circular permutations represented by ABC and its 'reverse' (back to forth) CBA. Finally a circular permutation distinguishes between 'back' and 'forth' or clockwise and counter clockwise. If there is not difference ABC and ACB are also the same Thus for the general case we have $(n-1)!/2$ circular full-duplex permutations. ■

Exercise 272.
How many 5 digit number can be formed using the digits 1,1,2,3,3?

Solution.
$5!/(2!2!)$ is an obvious answer. ■

Exercise 273.
How many 5 digit number can be formed using the digits 1,1,0,3,3?

Solution.
$5!/(2!2!)$ is an obvious answer but for a 0 at the front it becomes a 4-digit number. We need to subtract these from the total: they are $4!/2!2!$. Thus $5!/(2!2!) - 4!/2!2!$. ■

Exercise 274.
Find the 25-th permutation in lexicographic order of the elements of the set $\{1,2,3,4,5\}$.

Solution.
The first $4! = 24$ permutations in lexicographic order start with an 1. Thus the 25-th starts with a 2. It is 2134. ■

Exercise 275.
Let $F(n)$, $n \geq 1$ be the number of permutations of the objects $1, \ldots, n$ so that $|i - f(i)| \leq 1$ for all $1 \leq i \leq n$.

Solution.
If object 1 is fixed i.e. $f(1) = 1$, then for the remaining $n-1$ objects and positions let the number of permutations be $F(n-1)$. (F for fixed but may be not.) If object 1 is not fixed then $f(1) = 2$. But then 2 can only move to position 1 and thus $f(2) = 1$, because any other object $3, \ldots, n$ would move 2 positions or more. The remaining permutation excluding object 1 and 2, and positions 1, and 2, is thus $F(n-2)$. By the principle of addition $F(n) = F(n-1) + F(n-2)$. Base cases $F(1) = 1$ and $F(2) = 2$. ■

5.6 Combinations: examples

Exercise 276.

Combinations. A selection of a set of k out of n objects is defined as a k-**combination** of those objects or just a combination of k objects. The k elements form a set that is order does not matter. We can call this a k-combination. The number of k-combinations $C(n,k)$ is thus

$$C(n,k) = P(n,k)/k! = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Solution.
The number of permutations k out of n is $P(n,k)$. For a given subset of k objects the $k!$ permutations of them should be counted once in a combination not $k!$. Thus $C(n,k) = P(n,k)/k!$.

$$C(n,k) = C(n,n-k)$$

$$C(n,k) = C(n-1,k) + C(n-1,k-1)$$

Function $C(n,k)$ has some important properties. $C(n,k) = C(n,n-k)$. Picking k out of n we are left with $n-k$ in the original pile of n objects. Thus picking k out of n also defines a combination of $n-k$ objects out of n.

For the next identity we can prove it by interpretation and by using induction (on $n-1$ objects twice for two different 'k's!). Pick one object say 5. Each combination among the $C(n,k)$ either contains 5 or does not contain 5. The number of combinations that do not contain 5 is $C(n-1,k)$: number of combinations to pick k objects out of the remaining $n-1$ after 5 is excluded. The number of combination that contain 5 is $C(n-1,k-1)$: one of the objects in the k combination is 5 and thus we need to fill the remaining $k-1$ positions from the remaining $n-1$ objects and we have $C(n-1,k-1)$ ways to do so. Some people refer to the identity as Pascal's identity (lemma).
Another way to prove it is through permutations. The number of k-permutations is $P(n,k)$. But $k!$ of those permutations map to the same subset. Thus the number of distinct subsets, the number of k-combinations is $P(n,k)/k! = n!/((n-k)!k!)$. The latter is $\binom{n}{k}$. ■

Exercise 277.
What is the number of 5-bit sequences with 2 ones and 3 zeroes?

Solution.
$\binom{5}{2} = \binom{5}{3}$. (You select the two positions that will be set to one; the other three will be set to zero.) ■

Exercise 278.
(a) What is the number of combinations of n objects out of n?
(b) What is the number of combinations of one object out of n?

Solution.
(a) Number of combinations of n objects out of n is $C(n,n) = P(n,n)/n! = n!/n! = 1$ obviously.
(b) Number of combinations of one object out of n is $C(n,1) = n$. ■

Exercise 279.
Let A,B,C be three objects. What is the number of combinations of 2 objects out of 3 ?

Solution.
Let A,B,C be three objects. We have found previously that the following 2-permutations of those three objects are available: AB, BA, AC, CA, BC, CB From these AB and BA give rise to combination $\{A,B\}$, and like wise AC, CA to $\{A,C\}$ and BC, CB to $\{C,B\}$. Thuse we have three combinations. Therefore $P(3,2)/2! = 3$ is the number of combinations of 2 objects out of 3. ■

Exercise 280.

Generate all two-subsets of the n integers $1, 2, \ldots, n$. A two-subset is a subset of cardinality two. All generated subsets must be distinct.
Show that your algorithm is (time) optimal.

Solution.

```
for(i=1; i< n ; i++) {
  for(j=i+1; j < n+1; j++) {
    print(i,j);
  }
}
```

The running time i.e. the number of iteration steps is

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = n(n-1)/2.$$

Given that there are $\binom{n}{2}$ two-subsets the result follows. ■

Exercise 281.
There are three piles of 5 20-dollar bills, 6 10-dollar bills, and 7 5-dollar bills. We are to pick 3 20-dollar bills, 4 10-dollar bills, and 5 5-dollar bills. What is the number of ways to pick them up?

Solution.
The number of ways they can be picked up is

$$\binom{5}{3}\binom{6}{4}\binom{7}{5}$$

■

Exercise 282.
In a byte, we have a bit sequence consisting of 3 0s and thus 5 1s. What is the number of such bit sequences that contain 3 0s and 5 1s?

Solution.
The number of such bit sequences is $\binom{8}{3} = \binom{8}{5}$. ■

Exercise 283.
What is the number of k subsets of $2n$ objects $1,2,\ldots,2n$ where no objects are adjacent in a k subset? (That is you will never have $1,2$ in the subset not $4,5$ etc.)

Solution.
Represent a subset of k elements with a binary sequence of length $2n$ that has exactly k ones. Moreover there are no two consecutive ones. We use to count the binary sequences the previous result. THe binary sequence has $2n-k$ 0s and thus k 1s. Put down the $2n-k$ 0s; they create among themselves $2n-k+1$ bins (or buckets). Number the buckets $1,2,3,\ldots 2n-k+1$. The k ones can be placed into those buckets, one 1 per bucket. The number of ways this placement can take is $\binom{2n-k+1}{k}$. ■

Exercise 284.
We have a candy shop with 4 types of candies. We have vanilla, chocolate, strawberry, and banana candies (V, C, S, B). How many ways can we pick 5 candies? We could pick VVVVV, VVCSB, SSSSS, BBBCC, and so on. Counting is not easy.

Solution.
We would list the candies namelessly representing one instance with 1 on a line first V, then C, then S, then B In betwen there would a bar used as a separator Thus $11|1|1|1$ represents VVCSB and $11111|||$ represents VVVVV, and $|11||111$ representing CCBBB. Thus we have 5 one (number of candies) and 3 separators (number of candy types minus -1). The total string length (of ones and vertical bars) is thus $5+3=8$. We need to figure out the number of ways we can place the 3 bars, in fact turning 8 ones into 5 ones and 3 bars. This is $C(8,5)=8!/5!3!=56$. In general if the number of candy types is T and we pick D of them, the number of ways we can pick them is $C(D+T-1,D)$. ■

Exercise 285.
(a) An urn contains 10 balls labeled 1 through 10. In how many ways can we place those 10 balls into three (labeled) boxes, with 4 balls into box 1, 4 balls into box 2 and, 2 balls into box 3?
(b) An urn contains 10 balls labeled 1 through 10. In how many ways can we place those 10 balls into three (unlabeled) boxes, with two boxes containing 4 balls and one box containign two balls?

Solution.
(a)

$$\binom{10}{4}\binom{6}{4}\binom{2}{2} = \frac{10!}{4!4!2!}.$$

(b) Start with the previous problem and thus first obtain the answer

$$\frac{10!}{4!4!2!}.$$

But the two boxes of 4 balls can be swapped in order as they are not labeled. Thus if have n_m boxes of size m, we need to divide the result by $n_m!$. The final result, since we have 2 boxes of capacity 4, and 1 box of capacity 2, is then

$$\frac{10!}{4!4!2!} \cdot \frac{1}{2!1!}.$$

■

5.7 Ordered items : examples

Exercise 286.
Suppose a shop has m kinds of postcards and you have n friends.
(i) In how many ways can you send cards so that every friend gets one card?
(ii) In how many ways can you send cards so that every friend gets at most one card?
(iii) If you buy one of each kind of card, in how many ways can you send them to friends (a friend may get no cards, or more than one card)?
(iv) When was the last time you mailed a postcard to a friend?

Solution.
(i) For each of the n friends there are m choices. The answer is m^n.
(ii) For each of the n friends there is now one more choice (the possibility of not sending a card). The answer now becomes $(m+1)^n$.
(iii) For each of the m cards there are n choices. The answer is n^m.
(iv) No comment. ■

Exercise 287.
There are m different postcards and n diferent friends. (a) What is the number of ways of sending cards to friends? (b) What is the number of ways of sending distinct cards to friends?

Solution.
There are m different postcards and n diferent friends. (a) The number of ways of sending cards to friends is m^n or $(m+1)^n$ if one option is to NOT send a card. (b) The number of ways of sending distinct cards to friends is $m(m-1)\dots(m-n+1) = m!/(m-n)!$.

■

Exercise 288.
Find the (non-negative) integer solutions for

$$x_1 + x_2 + \ldots + x_n = k$$

Solution.
(See review.) Write down k o markers and arrange $n-1$ | markers around them. The answer is

$$\binom{k+n-1}{n-1} = \binom{k+n-1}{k}.$$

What if we looking for solutions that positive integer i.e. $x_i > 0$? Reduce this to a $y_i = x_i - 1$ problem! ■

Exercise 289.
(i) Find the number of solutions $(x_1, x_2, \ldots, x_n)$, where each $x_i \geq 0$ is an integer, to the equation

$$x_1 + x_2 + x_3 + \cdots + x_n = r$$

(ii) Find the number of solutions when each $x_i > 0$.

Solution.
(i) Let $x_1 + x_2 + \ldots + x_n = r$ where $x_i \geq 0 \ \ \forall i$. Put $y_i = 1 + x_i$. Then $y_i \geq 1 \ \forall i$. Then we get that $y_1 + y_2 + \ldots + y_n = n + r$. Therefore the number of solutions in positive y_i's is equal to the number of solutions in non-negative x_i. If we take $n + r$ dots and place $n - 1$ marks in the $n + r - 1$ spaces between the dots, we may take y_1 to be the number of the first set of dots, y_2 as the number of the second set and so on. Therefore the number of solutions to the original problem is $\binom{n+r-1}{n-1}$. Another way to see this problem is to count the number of ways r indistinct balls can be put into n distinct bins.

(ii) If instead of having $n + r$ in equation $y_1 + y_2 + \ldots + y_n = n + r$, we have r we get the number of solutions for the second part (assuming $r \geq n$) i.e. $\binom{r-1}{n-1}$. ■

5.8 Combinatorial interpretation : examples

Exercise 290.

The factorial function is defined as follows: $f(n) = n!$ and reads "n factorial". Note that $f(0) = 0! = 1$ and $f(1) = 1! = 1$. Moreover $f(n) = n \cdot f(n-1) = n \cdot (n-1)!$. For $n > 10$, we have $n! \approx \sqrt{2\pi n}(n/e)^n$ A better approximation is

$$(n/e)^n\sqrt{2\pi n}e^{1/(12n+1)} \leq n! \leq (n/e)^n\sqrt{2\pi n}e^{1/(12n)}.$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

The following is immediately derived if we rearrange (swap) in the denominator above $k!$ and $(n-k)!$.

$$\binom{n}{k} = \binom{n}{n-k},$$

Prove it.
Moreover,

$$\binom{-n}{k} = (-1)^k\binom{n+k-1}{k}$$

and

$$\binom{-1/2}{k} = \frac{(-1)^k}{4^k}\binom{2k}{k}$$

Solution.
One can prove such a fact directly or by combinatorial intepretation. Suppose we have n distinct objects. Suppose we label the objects $1,\ldots,n$. We can ask in how many ways can we pick k distinct object out of n and the answer is $\binom{n}{k}$. We have n ways to pick the first object, $n-1$ ways to pick the second object (having made the selection of the first), and so on, $(n-k+1)$ to pick the k-th object. However with this method we count the same collection $k!$ times. Whether we pick say three objects in this sequence $2,1,3$ or that sequence $3,2,1$ or etc is immaterial as the outcome in all 6 available possibilities draws to picking objects 1,2, and 3. Thus we need to divide the product $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ with $k!$. The result follows if we multiply numerator and denominator with $(n-k)!$.

To prove the left hand side is equal to the right hand side then becomes trivial. Picking k of n objects is equivalent to assigning a label 'Picked' to k out of n objects. This can be done in $\binom{n}{k}$ by the previous argument. However an alternative is to assign to $n-k$ out of the n objects the lable 'Unpicked'. This can be done in $\binom{n}{n-k}$. But then the objects without an 'Unpicked' label are assigned the 'Picked' label. Thus the number of way to pick k objects out of n is equal to the number of ways to unpick $n-k$ objects out of n (and then the not-unpicked objects are subsequently picked). ■

Exercise 291.

Using mathematical induction the following can be proven. It is the general form of the Binomial Theorem. (It was proved in an earlier chapter.)

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Then for $a = b = 1$ we have

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n,$$

For $a = -1$ and $b = -x$ and $n = -n$ we have

$$(-1-x)^{-n} = \sum_{k=0}^{n} \binom{-n}{k} (-1)^k (-x)^{-n-k}.$$

By simple math manipulations the following can then be proven. This is the basis used in the Pascal Triangle to determine the coefficient of $a^k b^{n-k}$ i.e. $C(n,k)$ from the coefficients of $C(n-1,k)$ and $C(n-1,k-1)$. Note $C(n,k) = \binom{n}{k}$.

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

Prove it.

Solution.

Let us call the objects $1,\ldots,k$. Let us consider object 1. When we choose k out of n objects 1 can be picked or not.

Case 1 (object 1 is not picked). Then the number of ways of picking k objects out of n is to not-pick 1 and pick k objects out of the remaining $n-1$ objects (the remaining ones are objects $2,\ldots,n$). The latter can be done in $\binom{n-1}{k}$ ways.

Case 2 (object 1 is picked). Then the number of ways of picking k objects out of n is to pick 1 and then pick $k-1$ objects out of the remaining $n-1$ ones (objects $2,\ldots,n$). The latter can be done in $\binom{n-1}{k-1}$ ways. By the addition principle the results follows. ■

Exercise 292.
Show that

$$1 \cdot 1! + 2 \cdot 2! + \ldots i \cdot i! + \ldots n \cdot n! = (n+1)! - 1$$

by combinatorial interpretation.

Solution.
The right hand side gives a hint. It is the number of permutations of $(n+1)!$ objects excluding the identity permutation. Then $i \cdot i!$ is the number of permutations where $p_{i+2} \ldots p_{n+1}$ are in place but p_{i+1} is not. ■

Exercise 293.
Show that

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{k}+\ldots+\binom{n}{n}=2^n$$

by combinatorial interpretation.

Solution.
2^n maps to the number of subsets of a set of n distinct objects. Each $\binom{n}{k}$ gives the number of subsets with k out of the n objects. ■

Exercise 294.
Show that

$$\binom{n+m}{k} = \binom{n}{0}\binom{m}{k} + \binom{n}{1}\binom{m}{k-1} + \ldots + \binom{n}{i}\binom{m}{i-1} + \ldots + \binom{n}{k}\binom{m}{0}$$

by combinatorial interpretation.

Solution.
Two types of objects, n of type A and m of type B. Choose k out of the $n+m$ of them. ■

Exercise 295.
Show that

$$\binom{n}{m}\binom{m}{k} = \binom{n}{k}\binom{n-k}{m-k}$$

by combinatorial interpretation.

Solution.
Number of ways to split n objects into three groups of k, $m-k$ and $n-m$ objects. ■

Exercise 296.
Let p,q,r be non-negative integers with $r \leq p$ and $r \leq q$. Then

$$C(p+q,r) = \sum_{i=0}^{i=r} C(p,i) \cdot C(q,r-i).$$

Prove it by combinatorial interpretation.

Solution.
We prove by combinatorial interpretation. Say we have p red labeled objects, and q blue labeled objects. (Labels are distinct.) In how many combinations can we pick r out of the $p+q$ objects? In an r combination i objects are red and thus $r-i$ objects would be blue. Fill-in the details. The answer is $C(p+q,r)$. ■

Exercise 297.
By giving a combinatorial interpretation for the following sum deduce a simple expression for it:

$$\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\ldots+n\binom{n}{n}$$

Solution.
This sum counts the ways that one may take a subset of k out of n items, and then take another one item out of the k (for k from 1 to n). This is equivalent to selecting one item out of the n and then splitting the remaining $n-1$ into two arbitrary subsets. This may be done in $n2^{n-1}$ ways, giving us the more concise value for the sum. ■

Exercise 298.
By giving a combinatorial interpretation to each side prove that:

$$\left(\begin{array}{c} 2n \\ n \end{array} \right) = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right)^2$$

Solution.
The left hand side is the number of ways of dividing $2n$ objects into two groups of equal size. Another way to count that same quantity is to split the $2n$ objects into two halves and then count all the ways one can take k objects from the first half and $n-k$ from the second half, over all values of k from 0 to n. This method of counting tells us that

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k} \left(\begin{array}{c} n \\ n-k \end{array} \right) = \sum_{k=0}^{n} \binom{n}{k}^2$$

■

Exercise 299.
Show that for any $n > r$

$$\sum_{k=r}^{n}(-1)^k\binom{k}{r}\binom{n}{k}=0$$

Solution.
From combinatorial arguments we know that

$$\sum_{k=r}^{n}(-1)^k\binom{k}{r}\binom{n}{k} = \sum_{k=r}^{n}(-1)^k\binom{n}{r}\binom{n-r}{k-r}$$

(The inner product in both cases is the number of ways n objects can divided into piles of size r, $k-r$, and $n-k$). Thus it suffices to show that

$$\begin{aligned}\sum_{k=r}^{n}(-1)^k\binom{n-r}{k-r} &= \sum_{i=0}^{n-r}(-1)^{i+r}\binom{n-r}{i}\\ &= \pm\sum_{i=0}^{n-r}(-1)^{i}\binom{n-r}{i}\\ &= 0\end{aligned}$$

Note that

$$(1+x)^{(n-r)} = \sum_{i=0}^{n-r}x^i\binom{n-r}{i}$$

Thus,

$$\sum_{i=0}^{n-r}(-1)^{i}\binom{n-r}{i} = (1+(-1))^{(n-r)} = 0$$

■

Exercise 300.

Provide a combinatorial interpretation for each of the following two sums to deduce a simple expression for each of them.

(a)

$$\binom{n}{0}+2\binom{n}{1}+2^2\binom{n}{2}+\ldots+2^n\binom{n}{n},$$

(b)

$$\binom{n}{k}+\binom{n-1}{k-1}+\ldots+\binom{n-k}{0}.$$

Solution.

(a) Suppose we have n balls. The possible colorings of these balls with 3 colors (R,G,B) is 3^n. We express this number differently as follows. Pick k among the n balls, color the unpicked $n-k$ R, and the number of colorings of the remaining k with the two colors G, B is 2^k. Sum up for all values of k from 0 (all balls R colored) to n (no ball R colored) and it is straightforward that this sum is 3^k.

(b) We can choose k balls among $n+1$ balls in $\binom{n+1}{k}$ ways.

- If the first ball is not chosen we can choose k among the remaining n in $\binom{n}{k}$ ways.
- If the first ball is chosen but the second is not we can choose $k-1$ balls among the remaining $n-1$ ones in $\binom{n-1}{k-1}$ ways and these $k-1$ balls and the first one give the k chosen balls. We proceed similarly,
- If the fist ball is chosen, the second ball is chosen , the l-th ball is chosen but not the $l+1$-th one we can choose $k-l$ balls among the remaining ones in $\binom{n-l}{k-l}$ ways, etc,

Summing for all values of l from 0 to k we get that the sum counts the number of ways of choosing k among $n+1$ balls which is $\binom{n+1}{k}$. ■

Exercise 301.
Show the following identities by combinatorial interpretation:

$$\sum_{i=1}^{n} i\binom{n}{i}^2 = n\binom{2n-1}{n-1} \tag{5.1}$$

$$\sum_{r=k}^{n-k} \binom{r}{k}\binom{n-r}{k} = \binom{n+1}{2k+1} \tag{5.2}$$

$$\tag{5.3}$$

Solution.
Suppose we have $2n$ distinct objects, split evenly between two bins. The right hand side counts the number of ways we can select a single object from the first bin, and then divide the remaining objects into two groups of size $n-1$ and n. The left hand side counts the same quantity when viewed in the following way. For some $1 \leq i \leq n$, pick i objects from the first bin, and $n-i$ objects from the second bin. From among the i objects chosen from the first bin, pick the single element. The remaining $(i-1)+(n-i)$ objects chosen are the set of $n-1$, and the $2n-n$ objects not chosen are the set of n.
The right hand side counts the number of ways we can select an odd number of objects $(2k+1)$ from a set of $n+1$ distinct objects. The left hand side counts the same quantity when viewed in the following way. Suppose the $n+1$ objects are ordered. To pick the $2k+1$, consider separately the various possibilities for the "middle" element of the $2k+1$. This object must be at least the $k+1$'st and at most the $n+1-k$'th (since it must be preceded by and followed by k or more objects). If the middle object is at position i, the number of ways the remaining objects can be completed is $\binom{i-1}{k}$ (pick the first half) times $\binom{n-i}{k}$ (pick the second half). This is what is counted on the left hand side (where r is actually the position before the one of the middle element selected). ■

5.9 Combinatorial identities : examples

Exercise 302.
For what value of k is

$$\binom{n}{k}$$

maximized? Express it in terms of n.

Solution.
Given that

$$\binom{n}{k} = \binom{n}{n-k}$$

the max value is for $n-k=k$ i.e. $k=n/2$ for even n, and for $k=(n+1)/2$ and $k=(n-1)/2$ when n is odd. ■

Exercise 303.
Show that

$$k\binom{n}{k} = n\binom{n-1}{k-1}.$$

Solution.
One may view $k\binom{n}{k}$ as the number of combination of choosing a team of k players and appointing a person as the team leader. Another way to say this is picka leader first in n possible ways and then fill the remaining $k-1$ positions of the team from the remaining $n-1$ members. ■

Exercise 304.
Show for n,k such that $k \leq (n-1)/2$ we have the following

$$\binom{n}{k} \leq \binom{n}{k+1}$$

Solution.
From the previous problem $k \leq (n-1)/2$ yield $2k \leq n-1$ which leads to $k+1 \leq n-k$ thus leading to

$$\frac{1}{n-k} \leq \frac{1}{k+1}$$

Thus

$$\binom{n}{k} = \frac{n!}{k!(n-k-1)!} \cdot \frac{1}{n-k} \leq \frac{n!}{k!(n-k-1)!} \cdot \frac{1}{k+1} \leq \frac{n!}{(k+1)!(n-k-1)!} \leq \binom{n}{k+1}$$

■

Exercise 305.
Show for n,k such that $k \geq (n-1)/2$ we have the following

$$\binom{n}{k} \geq \binom{n}{k+1}$$

Solution.
From the previous problem $k \geq (n-1)/2$ yields $n-k-1 \leq n-(n-1)/2-1$ which is simpliefied to $n-k-1 \leq (n-1)/2$ which from the previous problem yields

$$\binom{n}{n-k-1} \leq \binom{n}{n-k}$$

Thus

$$\binom{n}{k+1} \leq \binom{n}{k}$$

and thus the result is proven. The conclusion of the two previous problem (including this one) is that the maximum of $\binom{n}{k}$ occurs for $k = \lfloor (n/2) \rfloor$ and is $\binom{n}{\lfloor (n/2) \rfloor}$. ■

Exercise 306.
Let n be a positive integer. Then for all x, y we have

$$(x+y)^n = \sum_{k=0}^{k=n} \binom{n}{k} x^k y^{n-k}.$$

Prove it.

Solution.

$$(x+y)^n = \overbrace{(x+y)}^{\text{1st term}} \cdot \overbrace{(x+y)}^{\text{2nd term}} \cdot \overbrace{(x+y)}^{\text{3rd term}} \cdot \ldots \cdot \overbrace{(x+y)}^{\text{n-th term}}.$$

In order to form $x^k y^{n-k}$ we multiple the x of any k of the n terms that each one of them is $x+y$ with the y of the remaining $n-k$ terms that each one of them is $x+y$. Thus if we use the x from the first term and all $n-1$ remaining ys, and x of the second term and all remaining ys and so one we have formed $x^1 y^{n-1}$ in $n = \binom{n}{1}$ possible ways.
Another way to interpret this is by string formation. Strings would represent multiplications. The terms are x or y. There is only one way to generate the all-x string of length n. Likewise for the all-y string of length n. But how many strings do we have that have k x and $n-k$ y? It is $C(n,k) = \binom{n}{k}$. ■

Exercise 307.
Show the following

$$\binom{n}{k-1}+\binom{n}{k}=\binom{n+1}{k},$$

for any $1 \leq k \leq n$.

Solution.

$$\begin{aligned}
\binom{n}{k-1}+\binom{n}{k} &= \frac{n!}{(k-1)!(n-k+1)!}+\frac{n!}{k!(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{1}{n-k+1}+\frac{1}{k}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\cdot\frac{n+1}{k(n-k+1)} \\
&= \frac{(n+1)!}{k!(n-k+1)!} \\
&= \binom{n+1}{k}
\end{aligned}$$

■

Exercise 308.
Evaluate the following series.

$$\binom{n}{k}+\binom{n-1}{k-1}+\binom{n-2}{k-2}+\cdots+\binom{n-k+1}{1}+\binom{n-k}{0}$$

Solution.
Using

$$\binom{n+1}{k}=\binom{n}{k}+\binom{n}{k-1}$$

and simultaneous induction on n,k , we get that

$$\binom{n}{k}+\binom{n-1}{k-1}+\binom{n-2}{k-2}+\cdots+\binom{n-k+1}{1}+\binom{n-k}{0}=\binom{n+1}{k}.$$

Another way to solve this problem is by combinatorial arguments. The number of ways to choose k elements out of $n+1$ is equal to the number of ways to choose k out of n when element 1 is not included in the set of k elements or to choose $k-1$ out of $n-1$ when 1 is included but 2 is not or to choose $k-2$ out of $n-2$ when 1,2 are included but 3 is not included or so on or choose 0 out of $n-k$ when $1,2,\ldots,k$ are all included in the set. ■

Exercise 309.
Evaluate the following series.

$$\binom{n}{0}+2\binom{n}{1}+2^2\binom{n}{2}+\cdots+2^n\binom{n}{n}$$

Solution.
From

$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}x^i y^{n-i},$$

if we set $x=2$,$y=1$ we obtain the answer: 3^n. ■

Exercise 310.
Find a closed form expression for

$$\sum_{i=0}^{n}\sum_{j=0}^{i} 2^i 2^j \binom{n}{i}\binom{i}{j}$$

Using calculus and generating functions, determine the value of the sum

$$\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\ldots+n\binom{n}{n}$$

Solution.
We show the answer is 7^n by applying the binomial theorem backwards.

$$\begin{aligned}
\sum_{i=0}^{n}\sum_{j=0}^{i} 2^i 2^j \binom{n}{i}\binom{i}{j} &= \sum_{i=0}^{n} 2^i \binom{n}{i} \sum_{j=0}^{i} 2^j \binom{i}{j} \\
&= \sum_{i=0}^{n} 2^i \binom{n}{i} (1+2)^i \\
&= \sum_{i=0}^{n} 6^i \binom{n}{i} \\
&= (1+6)^n
\end{aligned}$$

■

5.10 Grid examples

Exercise 311.

Grid climbing. We are on the cartesian $\mathbb{Z} \times \mathbb{Z}$ plane, and in particular at $(0,0)$. We want to climb to $(4,4)$. Each step we can take can be forward or up thus from (i,j) we can move to $(i,j+1)$ or $(i+1,j)$. In how many ways can we reach our destination?

Solution.

We need 8 steps 4 up steps and 4 right steps. Thus the problem can be simplified into a 8 long string of 4 U for up and 4 R for moving to the right. Possible solutions are $UUUURRRR$, $URURURUR$, $RURURURU$, and so one. Answer is $8!/4!4! = C(8,4)$. ■

Exercise 312.
We are on the cartesian $\mathbb{Z}\times\mathbb{Z}$ plane, and in particular at $(0,0)$. We want to climb to (n,n). In how many ways can we reach our destination?

Solution.
Generalizing the previous solution the answer is $C(2n,n) = (2n)!/n!n!$. ■

Exercise 313.

We are on the cartesian $\mathbb{Z}\times\mathbb{Z}$ plane, and in particular at $(0,0)$. We want to climb to (n,n). How many ways are there under the restriction that we never cross (go beyond the diagonal). (That is touch (i,i) are OK.)

Solution.

We have good and bad routes with the total $G_n+B_n=C(2n,n)$. If we have a bad route, we find the first time the diagonal is crossed on the way up. We then flip the directions of the path until we reach the end or we have a further violation, by turning from that point on an up move into a right move and a right move into an up move. If the cross was at (i,i) the new path has an edge from (i,i) to $(i+1,i)$. Up to that point we had $i+1$ up moves and i right moves. The remainder of the path has $n-i-1$ up and $n-i$ right moves. We flip the directions of the path after the crossing of the diagonal. The $n-i-1$ up become $n-i-1$ right moves, and the $n-i$ right moves become up moves. Adding to them the original $i+1$ up and i right moves we get a total $n-i+i+1=n+1$ up moves and $n-i-1+i=n-1$ right moves. The new path (the original from $(0,0)$ to $(i,i+1)$ and the flipped from $(i,i+1)$ to $(n-1,n+1)$). The total number of such paths is $C(2n,n-1)=C(2n,n+1)$. Each one of them maps to a bad path. The total number of paths from $(0,0)$ to (n,n) is $C(2n,n)$. The difference is the number of good paths i.e.

$$\begin{aligned}
C(2n,n)-C(2n,n-1) &= \binom{2n}{n}-\binom{2n}{n-1}\\
&= \frac{2n!}{n!n!}-\frac{2n!}{(n-1)!(n+1)!}\\
&= \frac{2n!}{n!(n-1)!}\left(\frac{1}{n}-\frac{1}{n+1}\right)\\
&= \frac{1}{n+1}C(2n,n)\\
&= C_n
\end{aligned}$$

■

Exercise 314.

(a) On the integer grid we want to move from the origin $(0,0)$ to (n,m) by moving only UP or RIGHT. In how many ways can we move? How about from $(0,0)$ to (n,n)?
(b) What if we add the extra constraint for the $(0,0)$ to (n,n) problem that the path never moves above the diagonal connecting the end-points?

Solution.
(a) Out of $n+m$ moves n must be RIGHT and m UP. So the number of ways is

$$C(n+m,n) = \binom{n+m}{n} = \binom{n+m}{m} = C(n+m,m).$$

For the subquestion we have

$$C(2n,n) = \binom{2n}{n}.$$

(b) In order to exceed the diagonal, just before moving above it with a UP (from now on U) move, the path was on it. Let k be the last time this happened i.e. point (k,k). Thus we determine first the no-up the diagonal paths from $(0,0)$ to (k,k). Then the path should move to $(k+1,k)$ and so on and the last move would be form $(n,n-1)$ to (n,n). Let the number of paths from $(0,0)$ to (k,k) be $A(k)$, then the number of paths from $(k+1,k)$ to $(n,n-1)$ is $A(n-k-1)$ noting that $(k+1,k)$ and $(n,n-1)$ is the diagonal of a notional square with endpoints $(k+1,k)$, (n,k), $(n,n-1)$, and $(k+1,n-1)$ and we use induction. Using the multiplication principle we deduce the number of paths from $(0,0)$ to (n,n) as follows.

$$A(n) = \sum_{k=0}^{n-1} A(k)A(n-k-1),$$

where $A(0) = 1$. Another way to prove the result is to subtract from all the routes $\binom{2n}{n}$ those that go above the diagonal. Any route that goes above the diagonal it hit for the FIRST time let us say the over diagonal at $(k,k+1)$. Consider the mirror path over the line connecting $(0,1)$ and $(n-1,n)$. This maps to a path from $(-1,1)$ to $(k,k+1)$. Conversely also such a path maps to a path originating from $(0,0$ that crosses and goes over the diagonal. So the number of up diagonal paths from $(0,0)$ to (n,n) is the number of paths from $(-1,1)$ to (n,n). Thus

$$\binom{n-1+n+1}{n-1} = \binom{2n}{n-1} = \binom{2n}{n+1}.$$

And therefore the number of paths in question from $(0,0)$ to (n,n) is

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}.$$

Utilizing generating functions

$$a(x) = A(0) + A(1)x + A(2)x^2 + \ldots,$$

we have

$$\begin{aligned} a(x) * a(x) &= (A(0) + A(1)x + A(2)x^2 + \ldots) * (A(0) + A(1)x + A(2)x^2 + \ldots) \\ &= \sum_{k=0}^{\infty} x^k (A(0)A(k) + A(1)A(k-1) + \ldots + A(k)A(0)) \\ &= \sum_{k=0}^{\infty} x^k A(k+1) \end{aligned}$$

Therefore $xa^2(x) = \sum_{k=0}^{\infty} x^{k+1} A(k+1) = a(x) - 1$. Then

$$xa^2(x) - a(x) + 1 = 0$$

Solving for $a(x)$ we have

$$a(x) = \frac{1 \pm \sqrt{1-4x}}{2x}.$$

Expanding for $-$, we have

$$a(x) = \frac{1 - \sqrt{1-4x}}{2x} = \frac{1}{2x}(1 - (1-4x)^{1/2}).$$

Therefore,

$$f(x) = 1 + \frac{4x}{2 \cdot 2!} + \frac{3 \cdot 4^2 x^2}{2 \cdot 2 \cdot 3!} + \cdots$$

Then

$$A(n) = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1) \cdot 4^n}{2^n (n+1)!} = \frac{2^n \cdot (2n)!}{(n+1)! \cdot 2^n \cdot n!} = \frac{1}{n+1} \binom{2n}{n}.$$

■

Exercise 315.

(i) One starts in the lower left corner of an $n \times n$ chess board and makes a series of moves, where each move is to go either one square to the right or one square up, so that one ends up in the top right corner. How many different paths are there?
(ii) How many different paths are there that avoid the deadly land mine planted at the square in row i and column j?

Solution.
(i) We can reach the top right corner from the lower left one after moving $(n-1)$ squares to the right and $(n-1)$ squares up. We can choose the upward movements in

$$C(2(n-1),(n-1)) = \left(\begin{array}{c} 2(n-1) \\ n-1 \end{array} \right) \tag{5.4}$$

different ways and thus also fix the rightward movements.

(ii) We count first the number of paths that cross the land mine. These are the number of paths we can follow from the lower left corner to the land mine times the number of paths from the land mine to the top right corner i.e. $C(i+j-2,i-1) \cdot C(2n-i-j,n-i)$ (reasoning same as in part (i)). Therefore the number of paths that avoid the land mine is equal to
$C(2(n-1),(n-1)) - C(i+j-2,i-1) \cdot C(2n-i-j,n-i)$. ■

Exercise 316.
Suppose from any point (i,j) on an infinite grid, one is allowed to make the moves $(i,j) \to (i+1, j+1)$ and $(i,j) \to (i+1, j-1)$. How many paths are there from $(0,0)$ to (m,n)?

Solution.
Suppose we make p "$\nearrow$" moves, and q "$\searrow$" moves. Since any move we make increases the x coordinate by one we must have $p+q=m$. Since the net change in the y coordinate is n, it must be that $p-q=n$. From adding these equations it follows that $p=\frac{m+n}{2}$, and hence the number of times each move is made fixed. But they can be taken in any order, so the answer is

$$\binom{p+q}{p} = \binom{m}{\frac{m+n}{2}}$$

(assuming m and n are both odd or even, otherwise it is impossible). ■

Exercise 317.

Suppose one can move on the integer grid from position $(x,y) \to (x+1,y+1)$ and from $(x,y) \to (x+1,y-1)$. Show, that the number of ways one can go from position $(0,0)$ to position (m,n) ($m,n>0$ and with $m+n$ even) without touching the *x-axis*(except in the beginning) is:

$$\binom{m-1}{\frac{m+n-2}{2}} - \binom{m-1}{\frac{m+n}{2}}$$

Solution.

Since every path from $(0,0)$ to (m,n) must not cross the x-axis at any intermediate point, the first possible move is to $(1,1)$ point. Take the image $(1,-1)$ of $(1,1)$ with respect to the x-axis. Take a path from $(1,1)$ to (m,n) which touches the x-axis for the first time at point $(x,0)$. Then every path from $(1,1)$ to $(x,0)$ above the x-axis maps 1-1 to a path (its reflection) from $(1,-1)$ to $(x,0)$ below the x-axis. The number of paths (unrestricted) from $(1,1)$ to (m,n) are $\binom{m-1}{\frac{m+n-2}{2}}$ while the number of paths from $(1,-1)$ to (m,n) (which cross the x-axis and each one of them corresponds to a path from $(1,1)$ to (m,n) that touches the x-axis) is $\binom{m-1}{\frac{m+n}{2}}$. Subtracting the second number from the first we get the number of paths from $(1,1)$ (and equivalently from $(0,0)$) to (m,n) which do not touch the x-axis.

■

5.11 Catalan number

Exercise 318.

The Catalan number of order $(n+1)$ is associated with the different number of full binary trees with $n+1$ leaves. Show the result.

Solution.

(Short diversion: a binary tree is a rooted and ordered tree. Rooted means a node is designated as the root. Ordered means the children of a node are labeled left vs rights. And it is a binary tree in the sense that a node can have 0, 1 or 2 children. A full binary tree is one where every node has 0 children, i.e. it is a leaf, or it has exactly two children.)

Thus $C(0)=1$, $C(1)=1$ and $C(2)=2$ (see examples below). For $C(n)$ the number of trees is the number of trees with a left subtree containing i leaves (and by induction the number of such subtrees is $C(i-1)$) and a right subtree containing $n-i$ leaves (and by induction the corresponding number is $C(n-i-1)$ for every i from 1 to n. (Note that we can not have an empty subtree as the root must have two children or the definition of a full binary tree gets violated.)

Some preliminary work. Let $C(x)$ be the o.g.f of c_n. Then

$$C(x)=\sum_{n=0}^{\infty} c_n x^n$$

We compute $C(x)*C(x)$.

$$C(x)*C(x)=\left(\sum_{n=0}^{\infty} c_n x^n\right)\cdot\left(\sum_{n=0}^{\infty} c_n x^n\right)=\sum_{i=0}^{i=n} c_i\cdot c_{n-i}x^n$$

Therefore $C(x)*C(x)$ is the o.g.f of $\sum_{i=0}^{i=n} c_i\cdot c_{n-i}$.

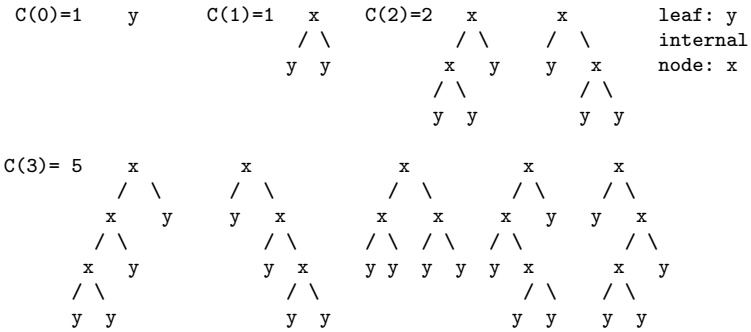
Thus $C(0)=1$ and for $n>0$ we have.

$$\begin{aligned}
C(n+1) &= \sum_{i=0}^{i=n} C(i)C(n-i)\\
\sum_{n=0}^{\infty} C(n+1)x^n &= \sum_{n=0}^{\infty}\sum_{i=1}^{i=n} C(i-1)C(n+1-i)x^n\\
\sum_{n=0}^{\infty} C(n+1)x^n &= C(x)\cdot C(x)\\
\sum_{n=0}^{\infty} C(n+1)x^{n+1} &= xC^2(x)\\
C(x)-1 &= xC^2(x)\\
xC^2(x)-C(x)+1 &= 0\\
C(x) &= \frac{1\pm\sqrt{1-4x}}{2x}
\end{aligned}$$

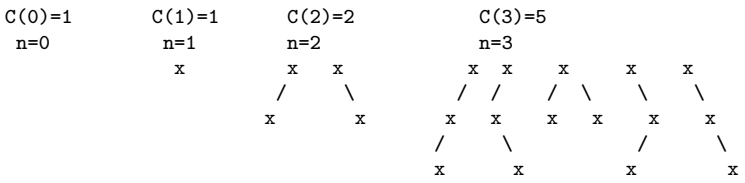
By the binomial theorem

$$(1-4x)^{1/2}=1-2\sum_{n=1}^{\infty}\binom{2n-2}{n-1}(-\frac{1}{4})^n\frac{(-4x)^n}{n}$$

Subtracting it from 1 and dividing by $2x$ the desired result is derived.



Let us examine now the number of non-isomorphic binary trees with n vertices.



Let $C(n)$ be the number of non-isomrphic binary trees with n nodes. If we have n vertices one becomes the root. The left subtree can have i generating by induction $C(i)$ non-isomorphic trees, and the right side $n-1-i$ with $C(n-i-1)$ trees. The total number is then $C(n) = \sum_{i=0}^{n-1} C(i)C(n-i-1)$. if we change n to $m+1$ we have $C(m+1) = \sum_{i=0}^{m} C(i)C(m-i)$. The if we change the name m back to n we have, $C(n+1) = \sum_{i=0}^{n} C(i)C(n-i)$. This is the first derivation in the previous proof that chows $C(n)$ is the n-th Catalan number. ■

5.12 Derangements of permutations

Exercise 319.
A derangement is a permutation where no object is in its initial position. Thus for ABC, the following are not DERANGEMENTS: ABC,ACB, BAC, CBA. However the following two permutations are derangements: BCA, CAB.
For a permutation with n objects (say $1,2,3,\ldots,n$) find the number of its derangements.

Solution.
Let A_i be the set of all permutations that fix i. Then $c(A_i) = (n-1)!$. Let $c(A_i \cap A_j) = (n-2)!$ i.e. all permutations that fix i and j. Thus we want to compute

$$c(A_1^{\complement} \cap A_2^{\complement} \cap \ldots \cap A_n^{\complement}) = n! - C(n,1)(n-1)! + C(n,2)(n-2)! + \ldots == n!(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}) \approx n!/e$$

■

Exercise 320.

Same problem. Show $D(n)$ the number of derangements of n objects is

$$D(n) = (n-1)(D(n-1)+D(n-2)),$$

where $n \geq 2$ with $D_0 = 1$ and $D_1 = 1$.

Solution.

Let, w.l.o.g., the n elements be $1,\ldots,n$. Element 1 has $n-1$ choices of placement. Say it is positioned at position 2. Then we examine two cases. Case 1 is that element 2 is in location 1. Then the elements $3,\ldots,n$ are to be placed in location $3,\ldots,n$ in $D(n-2)$ ways. Case 2 is that element 2 is not in location 1; then it must be in locations $3,\ldots,n$ and one of $3,\ldots,n$ in location 1. In other words we have $D(n-1)$ derangements. (2 does not go to 1, 3 does not go to 3, etc, n does not go to n.) The result follows. Note that $D(0) = 0$ and $D(1) = 0$, and $D(2) = 1$. Moreover $D(3) = 2$ and $D(4) = 9$. From

$$D(n) = (n-1)(D(n-1)+D(n-2)),$$

we define $E(n) = D(n) - nD(n-1)$ and obtain

$$D(n) - nD(n-1) = -(D(n-1) - (n-1)D(n-2)) \Rightarrow E(n) = -E(n-1).$$

Solving for $E(n) = (-1)^n$, we have the following.

$$D(n) = nD(n-1) + (-1)^n.$$

Note $D(0) = 1$. Then

$$D(n) = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{(-1)^n}{n!}\right).$$

■

Exercise 321.

Partial derangements. (a) Calculate the number of permutations of elements $1,\ldots,n$ in which k of the number are in their fixed positions and $n-k$ are not.
(b) Conclude then that

$$n! = \sum_k \binom{n}{k} D(k)$$

Solution.
(a) There are $\binom{n}{k}$ ways to pick k out of n objects that will be fixed. The remaining $n-k$ must form a derangement and this can be done $D(n-k)$ ways. So the total is

$$\binom{n}{k} D(n-k) = \binom{n}{n-k} D(n-k).$$

Any of the $n!$ permutations fixes k objects for some $k=0$ through $k=n$. Therefore,

$$n! = \sum_k \binom{n}{k} D(n-k) = \sum_k \binom{n}{n-k} D(n-k) = \sum_k \binom{n}{k} D(k).$$

■

Exercise 322.
What is the average number of fixed points in a random permutation?

Solution.
Utilizing the previous problem we obtain

$$\frac{1}{n!}\sum_{k=0}^{n} k\cdot\binom{n}{k}D(n-k)=\frac{1}{n!}\sum_{k=1}^{n} n\cdot\binom{n-1}{k-1}D(n-k)=\frac{1}{(n-1)!}\sum_{k=1}^{n}\binom{n-1}{n-k}D(n-k)=\frac{1}{(n-1)!}\sum_{k=0}^{n-1}\binom{n-1}{k}D(k)=\frac{1}{(n-1)!}(n-1)!=1.$$

■

5.13 Towers of Hanoi

Exercise 323.

The towers of Hanoi is a puzzle consisting in one form of three towers and n holed ringed discs. (each disc has a hole in its center so that it can slide through a tower). Each disc is of a different size. The discs may be moved from one tower to another tower provided that it is done one disc at a time and never can a disc be placed on top of a smaller disc. Initially, all n discs are in one tower.

(i) Formulate a recursive procedure to move the entire group to another tower using h_n individual moves, determined by the recurrence

$$\begin{aligned} h_i &= 2h_{i-1}+1 \\ h_1 &= 1. \end{aligned}$$

(ii) Can you improve this recurrence if there is a fourth tower?

Solution.
(i) Let the three towers be *input*, *output* , and *temp* . Procedure TOWERS shows how to legally move the towers from *input* to `output`.

```
 /* Discs are numbered 1 to n. */
 /* Operation move(i,j,k) moves disc  i from tower j to tower k */
 /* Precondition: assumes i has nothing on top of it            */
1.    Towers(input,output,temp,n) {
2.      if ( n==1 ) move(1,input,output);
3.      else {
4.             Towers(input,temp,output,n-1);
5.             move(n,input,output);
6.             Towers(temp,output,input,n-1)
7.      }
8.    }
end
```

If h_n gives the number of moves to move n discs in the algorithm above, then we have that $h_n = h_{n-1} + 1 + h_{n-1} = 2h_{n-1} + 1$, $h_1 = 1$, by way of lines 4-6. The base case $h_1 = 1$ is by way of line 2. The solution of this recurrence is $h_n = 2^n - 1$.
(ii) First, we show how a fourth tower helps. The simplest way to achieve this (let the four towers be *input, output, temp1, temp2*) is by moving recursively the top $n-2$ discs of the $input$ tower to say, $temp1$, then move disc $n-1$ to $temp2$, n to $output$, $n-1$ from $temp2$ to $output$ and recursively the $n-1$ discs from $temp1$ to $output$. The recurrence that describes the number of moves is the following: $b(n) = 2b(n-2) + 3$, $b(0) = 0$, $b(1) = 1$. A solution to this is the following

$$b(n) = \begin{cases} 2 \cdot 2^{\frac{n+1}{2}} - 3 & n \text{ odd} \\ 3(2^{n/2} - 1) & n \text{ even} \end{cases}$$

Based on this observation, we now get the following solution. Move the top $\frac{n}{2}$ discs (we assume n even, we can handle similarly the odd case) from the $input$ tower to $temp1$ using all 4 towers. Then move the $\frac{n}{2} - 1$ discs from $input$ to $temp2$ using only the $input$,$temp2$,$output$ towers as in part (i). Then move disc n from $input$ to $output$, then the $\frac{n}{2} - 1$ towers from $temp2$ to $output$ as in part (i) (using only 3 towers) and the $\frac{n}{2}$ from $temp1$ to $output$ as in part (ii) (recursively). We now give the recurrences that describe the number of moves in both the even and odd case.
For the even case we have:

$$c(n) = 2c(n/2) + 2a(n/2 - 1) + 1 = 2c(n/2) + 2(2^{n/2-1} - 1) + 1 = 2c(n/2) + 2^{n/2} - 1.$$

For the odd case we have:

$$c(n) = 2c(\frac{n+1}{2}) + 2^{(n-1)/2} - 1.$$

One can show that $c(n) \approx (1+\varepsilon)2^{n/2}$, where $\varepsilon \approx 2 \cdot 2^{-n/4}$. ■

5.14 Fibonacci related examples

Exercise 324.

How many n-digit binary 0-1 sequences have no adjacent 0's? Express the answer in terms of Fibonacci numbers.

Solution.
Let a_n be the number of binary strings of length n with no adjacent 0s. If the first digit is 1 we can have a_{n-1} strings for the other digit positions. If the first digit is 0 then the second digit must be 1 (otherwise we get two consecutive 0s) and we can have a_{n-2} strings for the rest $n-2$ digit positions. The recurrence we got is the Fibonacci recurrence shifted to the left one position i.e.

$$a_n = a_{n-1} + a_{n-2} \qquad a_1 = 2, \;\; a_2 = 3$$

and therefore the answer is f_{n+1} where f_n is the n-th Fibonacci number. (Note that several cases f_n might have defined initial values at f_0, f_1 or f_1, f_2.) ■

Exercise 325.

Show by combinatorial interpretation that the sequence

$$f_n = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} \binom{i}{n-i} \quad n = 0, 1, \ldots$$

is the Fibonacci sequence.

Solution.
The number of binary strings of length $n-1$ with no 2 consecutive 0s is f_n. We can also count them as follows. Let the number of 1s in the string be j. Then, we can put $n-1-j$ 0's in the $j+1$ slots formed by the 1s (one 0 per slot). Summing for all values of j we get (below we must have $j+1 \geq n-i-j$ so that no two 0s are consecutive, and the solution of this inequality gives the bounds for j):

$$\sum_{j=\lceil \frac{n}{2} \rceil - 1}^{n-1} \binom{j+1}{n-1-j} = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} \binom{i}{n-i}$$

This proves our claim. This sum is also the coefficient of x^n in $A(x) = \frac{1}{1-x-x^2} = \sum_i (x+x^2)^i$. ■

Exercise 326.

What is the number of n-bit sequences with no two consecutive ones?

Solution.
Fibonacci! Let f_n be the number of n bit sequences with no two consecutive ones. If the left-most bit is 0 the remaining $n-1$ bit can give rise to f_{n-1} $(n-1)$-bit sequences with no two consecutive ones by induction. If the left-most bit is a 1 the next second from left bit MUST be a zero, and the remaining $n-2$ bit can give rise to f_{n-2} $(n-2)$-bit sequences with no two consecutive ones by induction. The addition principle gives $f_n = f_{n-1} + f_{n-1}$. Beware! $f_1 = 2$ since 0, 1 are valid. Moreover $f_2 = 3$ since 00, 01, 10 are valid but not 11. The Fibonacci sequence is $F_n = F_{n-1} + F_{n-2}$, $F_0 = 0, F_1 = 1$ and $F_2 = 1, F_3 = 2, F_4 = 3$. Thus $f_n = F_{n+2}$! ■

5.15 Miscellaneous examples

Exercise 327.
In a unit square 5 arbitrary points are picked in its interior. Show that there exist two points among the five whose distance is less than $\sqrt{2}/2$.

Solution.
Split the square into four squares S_1, S_2, S_3, S_4 labeled counter-clockwise. Among the 5 points we have at least 2 falling into one of the squares. The max distance between two points is the counter diagonal distance: $\sqrt{(1/2)^2+(1/2)^2} = \sqrt{2}/2$, as needed. ■

Exercise 328.
For an n-sided polygon find the number of diagonals.

Solution.
One way is to observe that for a given point b, the two points say a, c on either side of b are connected with it with and edge. Thus there are $n-3$ possibilities for a diagonal to the remaining $n-3$ points. We do so for each of n points and thus the diagonal total is $n(n-3)$. This total counts diagonal (a,d) twice once as a diagonal for a and once for d. Thus the unique diagonal total is $n(n-3)/2$.
Another way is to observe that for n points the number of line segments is $n(n-1)/2$, since every point is connected with each one of the remaining $n-1$ points and the total is divided 2 as before. From this potential diagonal count we subtract the n edges of the polygon. Done. ■

Exercise 329.
How many integers $0 \ldots 9999$ do not contain 9?

Solution.
$(10^4 - 1) - (9^4 - 1) = 10^4 - 9^4$. ■

Exercise 330.
Number of n-bit sequences that have even number of 0s?
Number of n-bit sequences that have odd number of 0s?
Number of n-bit sequences that have even number of 1s?
Number of n-bit sequences that have odd number of 1s?

Solution.
The answer is 2^{n-1}. ■

Exercise 331.
Let $A = \{a,b,c,d\}$ and $B = \{1,2,3,4,5,6\}$. Find the number of functions $f : A \to B$ that are not 1-1.

Solution.
The number of functions from A to B is at $6 \cdot 6 \cdot 6 \cdot 6 = 6^4$. The ones that are 1-1 there are $6 \cdot 5 \cdot 4 \cdot 3 = 360$ Subtract and you have the answer. ■

Exercise 332.
How many sequences are there of 3 not necessarily distinct integers $x,\ y,\ z$ such that $1 \leq x,y,z \leq 100$ and the sum of them is divisible by 4.

Solution.
The first two numbers x and y may be chosen arbitrarily from the 100^2 possibilities. For any such choice, there are exactly $100/4 = 25$ possible z values that will force the sum to be a multiple of four. Thus there are $(100)(100)(25) = 250,000$ possible sequences. ■

Exercise 333.
(a) How many non-negative (consecutive) integers can we represent with 4 bits in binary? How many with n bits? How many with $\lg n$ bits?
(b) How many bits do we need to represent 7? How about 8? How about 15?
(c) How many bits do we need to represent $2^n - 1$? How many for 2^n?
(d) Assume n is a power of two for this question. How many bits do we need to represent all unsigned integers between 0 and $n-1$ in fixed width ? between 0 and $n^5 - 1$? (Such integers that will have fewer bit than the maximum $n^5 - 1$ will get leading zeroes.) Between 0 and $2^n - 1$? Between 0 and $n^n - 1$?

Solution.
(a) With 4bit we can represent $2^4 = 16$. With n bit, 2^n. With $\lg n$, we can $2^{\lg n} = n$.
(b) 7=0b111 i.e. 3 bits. 8=0b1000 i.e. 4 bits. 15=0b1111 i.e. 4 bits.
(c) $2^n - 1$ can be represented by n bits all of which are 1. 2^n can be represented by $n+1$ bits, an 1 followed by n 0s. (d) All integers between 0 and $n-1$ can be represented minimally with $\lg(n-1+1) = \lg n$ bits by part (c) above. All integers between 0 and $n^5 - 1$ can be represented with $\lg(n^5 - 1 + 1) = \lg n^5 = 5\lg n$ bits. All integers between 0 and $2^n - 1$ can be represented with $\lg(2^n - 1 + 1) = \lg 2^n = n$ bits. All integers between 0 and $n^n - 1$ can be represented with $\lg(n^n - 1 + 1) = \lg n^n = n \lg n$ bits. ■

Exercise 334.

(a) How many bit do we need to minimally represent $A = 2^{m-1}$? How many for $B = 2^m - 1$? Minimally means without leading zeroes. How many bit do we need to represent all unsigned integers between 2^{m-1} and $2^m - 1$? Express the answer as a function of m.

(b) Use the result from (a) to show that every positive integer L can be represented using $\lceil \lg(L+1) \rceil$ bit.

Solution.
(a) $A = 2^{m-1}$ is an 1 followed by $m-1$ zeroes. $B = 2^m - 1$ is a sequence of m 1's.
So both integers need m bit to be represented. If the extremes $A = 2^{m-1}$ and $B = 2^m - 1$ need m bit, so do all integers L between A and B such that $A \leq L \leq B$.
(b) Every positive integer L can be sandwitched between an A, and a B of part (a), for some integer value m. This means m bit are needed to represent (minimally) L. Thus

$$\begin{aligned} 2^{m-1} &\leq L \leq 2^m - 1 \\ 2^{m-1} + 1 &\leq L + 1 \leq 2^m \\ 2^{m-1} &< L + 1 \leq 2^m \\ \lg(2^{m-1}) &< \lg(L+1) \leq \lg(2^m) \\ m - 1 &< \lg(L+1) \leq m \end{aligned}$$

Bear in mind that $\lg(L+1) \leq \lceil \lg(L+1) \rceil$. The term $\lg(L+1)$ is strictly greater than $m-1$ and at most m. Its ceiling $\lceil \lg(L+1) \rceil$, which is an integer, can not thus be $m-1$, because the smaller $\lg(L+1)$ is greater than $m-1$, and therefore $\lceil \lg(L+1) \rceil$ must be the next greater integer, i.e. m. Therefore, $m = \lceil \lg(L+1) \rceil$. One can use the same method, symmetrically on the right inequality, to also show that $m = \lfloor \lg L \rfloor + 1$.

$$\begin{aligned} 2^{m-1} &\leq L \leq 2^m - 1 \\ 2^{m-1} &\leq L < 2^m \\ m - 1 &\leq \lg L < m \\ m - 1 &\leq \lfloor \lg L \rfloor \leq \lg L < m \end{aligned}$$

Therefore, $\lfloor \lg L \rfloor = m - 1$ (as it cannot be m), and solving for m we find that $m = \lfloor \lg L \rfloor + 1 = \lceil \lg(L+1) \rceil$. ■

Exercise 335.

How many 3-char strings have a letter (character) repeated if the allowable characters (letters) are $\{a,b,c,d\}$.

Solution.
We want to find the cardinality of A. Sometimes A is difficult to compute. A better approach is find U and $A^{\complement}$ and use $c(A) = c(U) - c(A^{\complement})$. For a $---$ there are four character candidates for the first $-$, and so on for the second, third $-$. This defines U and $c(U)$. There are 4^3 strings of three characters from the set. This means $c(U) = 4^3$.
Now let us find out the number of strings that DO NOT contain a duplicate character (i.e. aab is not allowed.) This way we will establish $A^{\complement}$ and $c(A^{\complement})$. We have 4 choices for the first $-$, but then 3 choices (remaining ones) for the second $-$ and just two choice (the unused ones) for the third $-$, for a total of $4 \cdot 3 \cdot 2 = 24$. Thus $c(A^{\complement}) = 24$.
A simple subtraction shows $c(A) = c(U) - c(A^{\complement}) = 4^3 - 24 = 40$. ■

Exercise 336.
Show that $(n!)!$ is divided by $(n!)^{(n-1)!}$.

Solution.
Consider $n = 3$. Show $6!$ is divided by 6^2. Consider $n = 4$. Show $24!$ is divided by 24^6. If we have $n!$ objects and we divide them into $(n-1)!$ sets of size n each set being of the same kind completes the proof. ■

Exercise 337.

Password Insecurity. A computer has a weird password policy. A password is 7 or 8 characters. A character is lower-case letter or a digit. Each password must have at least one digit. What is the total number of passwords for an attack vector?

Solution.
Let A_7 be the number of 7 char passwords. Likewise A_8. The number of 7 char passwords that are all lower case is 26^7 and these are impermissible (forbidden). Thus $A_7 = 36^7 - 26^7$. Likewise $A_8 = 36^8 - 26^8$ and thus the answer is $A_7 + A_8$. ■

Exercise 338.

What is the number of zeroes in $1000!$? (Yes it is 1000 factorial.)

Solution.
1. Since $2 \cdot 5$ is 10, Let find the multiples of 5. Each one contributes with a even number a 10. There are $1000/5 = 200$ such multiples.
2. A multiple of 25 contributes one more 10 since $25 = 5 * 5$. There are $1000/25 = 40$ of them.
3. A multiple of 125 contributes one more 10 since $125 = 5 * 5 * 5$. There are $1000/125 = 8$ of them.
4. Is the answer $200 + 40 + 8$?
We have missed 625!
5. Since $625 = 5 * 5 * 5 * 5$ we have one more 5 contributed by 625.
6. The total is 49.
Conclusion: Think! ■

Exercise 339.
How many distinct seven digit integers can be constructed by permuting the digits $1,1,1,2,2,3,4$?

Solution.
Applying the formula for permutations (repetitions allowed), we get

$$\frac{7!}{3!\ 2!} = 420$$

■

Exercise 340.

(a) How many n bit integers do we have.
(b) What is the sum of the values of those n-bit integers?
(c) Consider all three bit integers. What is the such of their values?

Solution.
(a) This is a choose n object from a choice of two (0 and 1) to form a n-bit binary sequence. We have 2^n binary sequences of length n.
(b) The righmost, second from right, etc bit, is 0 half of the times (2^{n-1}) and 1 as well. The sum of the values for such digts at a given position is thus 2^{n-1}. The sum of the 2^n integers is thus

$$2^{n-1} + 2 \cdot 2^{n-1} + \ldots + 2^{n-1} \cdot 2^{n-1} = 2^{n-1}\left(2^n - 1\right).$$

(c) The sum $0+1+\ldots+7$ is 28. Substituting 3 for n in

$$2^{n-1}\left(2^n - 1\right) = 2^{3-1}\left(2^3 - 1\right) = 4 \cdot 7 = 28,$$

yields the same answer, obviously. ■

Exercise 341.
Calculate the number of trailing 0's in the fully expanded decimal representation of $200!$

Solution.
We need to find the largest power of 10 that divides $200!$. Since $10 = 2 \cdot 5$, we only need find the largest power of 5 that divides $200!$, since this is smaller than the corresponding largest power of 2.
The largest power of 5 that divides $200!$ is:

$$\lfloor \frac{200}{5} \rfloor + \lfloor \frac{200}{5^2} \rfloor + \lfloor \frac{200}{5^3} \rfloor = 40 + 8 + 1 = 49$$

The first term in the above formula counts the multiples of 5. The second one counts the multiples of 25 which contribute one more power of 5 each, and the third one counts multiples of 125. ■

Exercise 342.

How many ways can we put in a line 10 boys and 5 girls, so that no two girls are next to each other.

Solution.
First put the ten boys in a line. There are eleven spots where a girl may be inserted, but only one girl may be placed in any spot. Thus the answer is the number of ways five objects can be selected out of eleven, which is $\binom{11}{5} = 462$. If you believe that the boys and girls are distinct, then multiply that quantity by $10!5!$ to account for their various permutations. ■

Exercise 343.

How many ways can $3r$ balls be selected from $2r$ red balls, $2r$ blue balls, and $2r$ white balls?

Solution.
There are two (basically equivalent) ways of dealing with this problem: Without generating functions.
Break this into two disjoint cases. Either at least r red balls are chosen (in which case the number of blue balls must be between 0 and $3r$ minus the number of red balls), or fewer than r red balls are chosen (in which case the number of blue balls must be between r minus the number of red balls and $2r$). Once the number of red and blue balls are decided, the number of white balls is uniquely determined. The answer is therefore

$$\begin{aligned}
\sum_{i=r}^{2r}\sum_{j=0}^{3r-i} 1 + \sum_{i=0}^{r-1}\sum_{j=r-i}^{2r} 1 &= \sum_{i=r}^{2r}(3r-i+1) + \sum_{i=0}^{r-1}(r+i+1) \\
&= \sum_{i=1}^{r+1}(2r-i+2) + \sum_{i=1}^{r}(r+i) \\
&= r+1+\sum_{i=1}^{r}(2r-i+2) + \sum_{i=1}^{r}(r+i) \\
&= r+1+\sum_{i=1}^{r}(3r+2) \\
&= r+1+r(3r+2) \\
&= 3r^2+3r+1
\end{aligned}$$

■

Exercise 344.
(a) How many n bit 0-1 sequences are there?
(b) How many n bit 0-1 sequences are there with 5 ones?
(c) How many 10 bit 0-1 sequences are there with more ones than zeroes?

Solution.
(a) The answer is 2^n.
(b) $\binom{n}{5}$.
(c) The number of sequences with equal number of 0 and 1 is by (b) $\binom{n}{5}$. The total number of sequences of length 10 is 2^{10}. The number of length 10 sequences of unequal number of 0s and 1s is $2^{10} - \binom{10}{5}$. Half of them have more zeroes than ones, half fewer zeroes than ones. So the answer is $2^{10-1} - 0.5\binom{10}{5}$. Fill-in the details. ■

Exercise 345.
How many ways can one select k distinct integers from the set $\{1,2,\ldots,n\}$ so that no two are consecutive?

Solution.
This is equivalent to asking how many words of length n over $\{0,1\}$ have exactly k ones, no two of which are adjacent. We can count this by laying down the $n-k$ zeros, and then chosing k of the possible $n-k+1$ gaps in which to insert ones. Thus the answer is

$$\binom{n-k+1}{k}$$

■

Exercise 346.
Find a simple expression for the following sum:

$$\binom{n}{0}\binom{n}{k}-\binom{n}{1}\binom{n}{k-1}+\ldots+(-1)^i\binom{n}{i}\binom{n}{k-i}+\ldots+(-1)^k\binom{n}{k}\binom{n}{0}$$

Solution.
Observe that the i'th term in the summation is the product of the coefficient of x^i in $(1-x)^n$, times the coefficient of x^{k-i} in $(1+x)^n$. Therefore the entire sum is the coefficient of x^k in $(1-x)^n(1+x)^n=(1-x^2)^n$. This is the coefficient of $y^{k/2}$ in $(1-y)^n$, which is

$$\begin{array}{ll} 0 & \text{if } k \text{ is odd} \\ (-1)^{\frac{k}{2}}\binom{n}{\frac{k}{2}} & \text{if } k \text{ is even} \end{array}$$

■

Exercise 347.

How many positive integers less than 10^n (in the decimal scale) have their digits in non-decreasing order?

Solution.
We represent the problem as follows. We pick $n+9$ balls. Among these balls we will pick 9 of them, and these will become the separators that will split the other ones in 10, possibly empty, parts (the first part is on the left of the first separator, the tenth on the right of the last one, the other ones are between any two successive separators). We label the balls of part i with number $i-1$, and thus, we use numbers $0,\ldots,9$. If there's no ball in part i, then no label $i-1$ will be assigned to any ball. After we had chosen the separator balls, and labeled the other n balls, reading the labels from left to right we get a number whose digits are in non-decreasing order. It is easy to see that the number of ways we can choose the 9 separator balls among the $n+9$ ones is $\binom{n+9}{9}=\binom{n+9}{n}$, and among them, there are $\binom{n+9}{n}-1$ positive ones. ■

Exercise 348.

A coin is tossed $2n$ times. How many of the 2^{2n} possible outcomes have the number of heads and tails equalized for the first time after $2n$ tosses?
(a) Find the answer by relating this problem to the Catalan numbers.
(b) Give a second proof using combinatorial interpretation arguments (DO NOT use generating functions or recurrences).

Solution.
(a) The first question is easy. Let the two outcomes of a coin toss be H and T. Suppose that the outcome of the first coin toss is H. Then, we replace each H by a **(** and each T by a **)** and the connection to Catalan numbers is straightforward. If the first outcome is T, then we replace each T by a **(** and each H by a **)**. Therefore, the result is the sum of these two cases, i.e. twice the n-th Catalan number.
(b) We prove the result here by counting the number of ways we can move on the integer grid in some way. The integer grid is the real plane where we are only allowed to make either a $(x,y) \to (x+1,y+1)$, (a $+$) move, or a $(x,y) \to (x+1,y-1)$, (a $-$) move, where x,y are integers in both cases.
Question a.1. What is the number of paths we can traverse from $(0,0)$ to (n,m) $(n \geq m > 0)$ using $+/-$ moves?
Answer a.1 . Let each $+$ move be represented by an 1 and each $-$ move by a -1. Then we are counting the number of solutions to the equation $x_1 + \ldots + x_n = m$, where $x_i \in \{-1,+1\}$ $\forall i$. If p is the number of $+1$'s and q the number of -1's we get that we must have $p+q=n$ and $p-q=m$. If we solve this system we get $p=(n+m)/2$, $q=(n-m)/2$, and the number of ways we can go from $(0,0)$ to (n,m) thru $+$ or $-$ moves is simply $\binom{n}{p} = \binom{n}{q}$ (choose p 1's among the n choices and let the other $n-p=q$ choices be -1's). It is easy to generalize this result for any starting (say (a,b)) and final position (say (m,n)) (by finding the number of solutions in $\{-1,1\}$ of an appropriate equation such as $x_1 + \ldots + x_{m-a} = n-b$).
Question a.2. What is the number of paths we can traverse from $(0,0)$ to $(2n,0)$ by moves that lie only on the first quadrant (that is positive x and y coordinates) without touching the x-axis?
Answer a.2. Let us call this number Q. Q is equal to the number of such paths from $(1,1)$ to $(2n-1,1)$ since the first move must be a $+$ one, and the last move must be a $-$ one, otherwise we leave the first quadrant. Q is also the number of paths Q_1 we can go from $(1,1)$ to $(2n-1,1)$ unconditionally (that is we may touch or not the x-axis) minus the number of such paths Q_2, that always touch the x-axis. Q_1 can be found from the earlier question , and it is $Q_1 = \binom{2n-2}{n-1}$. Let's now try to find the number of paths that meet or cross the x-axis (that is Q_2). Let us consider such a path P from $(1,1)$ to $(2n-1,1)$, and let it touch the x-axis *for the first time* at point $(r,0)$. Let P_1 be the part of P between $(1,1)$ and $(r,0)$ and P_2 the part between $(r,0)$ and $(2n-1,1)$. We reflect P_1 with respect to the x-axis so that point $(1,1)$ maps to $(1,-1)$ while $(r,0)$ remains fixed. P_1 thus maps uniquely to a path of the fourth quadrant(positive x, negative y), call it P_1'. This path concatenated with P_2 gives a pathfrom $(1,-1)$ to $(2n-1,1)$ that is in 1-1 correspondence with P and crosses the x-axis. So, the number of paths Q_2 that touch or cross the x axis is just the number of all paths from $(1,-1)$ to $(2n-1,1)$, which is, from the earlier question , equal to $\binom{2n-2}{n}$. Subtracting, we get:

$$Q = Q_1 - Q_2 = \binom{2n-2}{n-1} - \binom{2n-2}{n} = \frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{n-1}{n}\frac{(2n-2)!}{(n-1)!(n-1)!} = \frac{1}{n}\binom{2n-2}{n-1}$$

We map every H outcome of the toin coss to a $+$ move, and every T to a $-$ move. Then Q above counts all outcomes in which the number of heads equalizes the number of tails for the first time after $2n$ tosses AND the first coin toss is a H. We must add to this the number of such outcomes when the fist coin toss is a T. We only then need to count the number of ways one can go from $(0,0)$ to $(2n,0)$ by $+/-$ moves with out leaving the fourth quadrant and without touching the x-axis, which, by symmetry from the latter question, is also Q. $2Q$ is thus the answer to the problem. ■

Exercise 349.

Generate all permutations of n objects e.g. $1,\ldots,n$.

Solution.

We use an inductive construction. If $p = \langle p_1,\ldots,p_{n-1}\rangle$ is a permutation of $n-1$ objects $1,ldots,n-1$, we can generate permutations of n objects by inserting n (the n-th object) in the n positions available of p (before p_1, between p_1 and p_2, and so on).

```
  GenPerm (n)
 1. i = 1 ;
 2. p=<1>
 3. for(i=2;i<=n;i++) {
 4.  for each q = < p1 , ... , pi-1 > {
 5.     r= <p1 ... pi-1 i>
 6.     for(j=i-1; j>=1 ; j--) {
 7.       r = r with j indexed value swapped with j+1;
 8.     }
 9.  }
10. }
```

■

Exercise 350.
Find the radix-9 representation of a denary i.e. radix-10 integer.

Solution.
Observe $10^n = (9+1)^n = \sum_k \binom{n}{k} 9^k$. ■

Exercise 351.
How many n-bit sequences are there with k 0s and $n-k$ 1s, $n-k \geq k-1$, with no two adjacent 0s?

Solution.
Place down the 1. They define $n-k+1$ bins into which the zeroes can be placed. (Bin 1 is one the left of the 1st one, Bin 2 between bin 1 and bin 2, and finally bin $n-k+1$ to the right of the $n-k$-th one.) No two 0s can be in the same bin. So the answer is $\binom{n-k+1}{k}$. ■

Exercise 352.
Find the determinant of the $n \times n$ matrix A and the permanent of the $n \times n$ matrix B

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 1 \end{pmatrix}$$

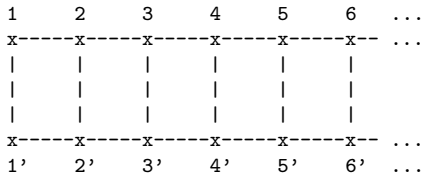
Solution.
(i) By decomposition of the first column, it is f_n, the order-n Fibnacci number.
(ii) Likewise. ■

Exercise 353.
Suppose we divide the numbers in the set $\{1,2,\ldots,n\}$ into l classes arbitrarily. Prove that if $n \geq e \cdot l!$ then there exists a class (among the l given classes) such that $x,\ y,\ z$ belong to this class and $x+y=z$.

Solution.
We are going to use a prior problem here. Let the l classes into which the integers from 1 to n were split be $C_1,\ldots,C_l$. Take the complete graph K_{n+1} vertices (we number its vertices with numbers from 1 to $n+1$) and color its edges with l colors as follows. We color edge $\{i,j\}$ with color m if the number $|\,i-j\,|$ is in the class C_m. Since the difference of any two vertices in K_{n+1} is at most n we can always color the edges in such a way. Since $n+1 \geq el!+1$, the graph has a monochromatic triangle, let it be (i,j,k,i) and let all edges of this triangle are colored m. This means that the numbers $|\,i-j\,|,\ |\,j-k\,|,\ |\,k-i\,|$ are all in set C_m. Let, without loss of generality, $i<j<k$. Then choose, $x=|\,i-j\,|$, $y=|\,j-k\,|$ and $z=|\,k-i\,|$. It is straightforward that $x+y=z$ and $x,y,z \in C_m$. Since $C_1,\ldots,C_l$ is an arbitrary splitting of the n numbers we are done. ■

Exercise 354.

Find the number of perfect matchings in the ladder graph shown below.



Solution.

Let $f(n)$ be the number of perfect matchings in the ladder graph of size n (with $2n$ vertices). If the edge $(1,1')$ is in the matching the other vertices can be matched in $f(n-1)$ ways. If edges $(1,2)$ and $(1',2')$ are in the matching the other vertices can be matched in $f(n-2)$ ways. That way we get the recurrence relation:

$$f(n) = f(n-1) + f(n-2) \quad n \geq 3\ , \quad f_1 = 1, \quad f_2 = 2$$

This is just the Fibonacci recurrence (for $n > 0$). ■

Exercise 355.
Consider arrangements in a line of a collection of r characters where one designated character appears k times, and each of the remaining $r-1$ characters appears once. Assuming $k \leq r$, how many such arrangements are there where no two occurrences of the designated character are consecutive?

Solution.
First consider arranging the $r-1$ characters that appear just once. This can be done in $(r-1)!$ ways. Now there are r slots (including the beginning and the end) where we can insert instances of the repeated character. We may place one instance in any such slot, but no more than one. Thus the total number of arrangements

$$(r-1)!\binom{r}{k}$$

■

Exercise 356.
In how many ways may you line up twenty (distinct) boys and ten (distinct) girls so that no two girls stand next to each other?

Solution.
There are 20! ways to permute the boys. However this is done, there are 21 slots between boys where one (but more) girl may be positioned (the front and back count as valid slots. We complete the process by choosing 10 of these slots, and ordering the girls within them. Thus the answer is

$$\binom{21}{10}10!20! = \frac{21!20!}{11!}$$

■

5.16 Distributions

Exercise 357.
Partitions. A partition is the allocation of indistinct objects into indistinct bins.
How many partitions does integer 5 have? For example one partition of five is $1,4$ since $1+4=5$. Since bins are indinstinct, this is the same as $4,1$.

Solution.
Integer 5 has 7 partitions.

$$\{1,1,1,1,1\},\{1,1,1,2\},\{1,2,2\},\{1,1,3\},\{2,3\},\{1,4\},\{5\}.$$

■

Exercise 358.

We have a set of 8 elements. Find the number n of ordered partitions of the set into three sets containing 3,2,3 elements respectively.

Solution.
Let the three sets be A_1, A_2, A_3. Since the set has 8 elements, A_1 can be formed in $C(8,3)$ ways. Then A_2 can be formed in $C(8-3,2) = C(5,2)$ ways and for A_3 we have $C(3,3) = 1$. Thus

$$n = C(8,3)C(5,2)C(3,3) = \frac{8!}{3!2!3!} = 2 \cdot 5 \cdot 7 \cdot 8 = 560.$$

If the partitions are unordered, $A_1A_2A_3$ but $A_3A_2A_1$ are the same; we then have

$$n/(2!) = 2 \cdot 5 \cdot 7 \cdot 8/2 = 280$$

■

Exercise 359.

(a) Partition a deck of cards among 4 people eachone getting 13. A deck of cards is partitioned into four set of 13 cards.

Solution.

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{(13!)^4}.$$

Or

$$(a1a2\ldots a12a13)(b1b2\ldots b12b13)(c1c2\ldots c12c13)(d1d2\ldots d12d13)$$

Set 52 place holders $a\ldots b\ldots c\ldots d\ldots$ and parentheses on a line. We have 52! to arrange the cards of the deck in each group. Within a group order does not matter, thus we divide by 13! for each group Total $52!/(13!)^4$. ■

Exercise 360.

Find the number of 3-ary numbers of length n with an odd number of 1s.

Solution.
Let a_n be the number of length n integers radix-3 with an odd number of 1s. If the first (say leftmost) digit is a 0, there a_{n-1} digits of length $n-1$ with odd number of 1s and thus also of length n with odd number of 1s. Likewise if the number starts with a 2. If the number starts with a 1 though, the remaining $n-1$ digits must have an even number of 1s. The number of numbers of length n with even number of ones is 3^{n-1} minus a_{n-1}. Thus

$$a_n = a_{n-1} + a_{n-1} + (3^{n-1} - a_{n-1}).$$

Therefore

$$a_n = a_{n-1} + 3^{n-1}.$$

The base case is $a_1 = 1$. Thus solving for a_n we have

$$a_n = 0.5(3^n - 1).$$

■

5.17 Stirling's number of the second order

Exercise 361.

Let $S(n,k)$, be the Stirling number of the second order. $SN(n,k)$ denotes the number of ways to partition an n-element set into exactly k partitions.

For boundary cases we have that $S(n,1) = 1$ and also $S(n,n) = 1$.

(a) Show that $S(3,2) = 3$.

(b) Then shown that for any $1 < k < n$ we have the following

$$S(n,k) = S(n-1,k-1) + kS(n-1,k).$$

Then conclude that the number of partition of a set of n elements.

$$\sum_{k=1}^{n} S(n,k),$$

Solution.

(a) Let the 3 element set be $\{1,2,3\}$. Then we have the following partitions.

$$\begin{aligned} \{1\} + \{2,3\} \quad &, \\ \{2\} + \{1,3\} \quad &, \\ \{3\} + \{1,2\} \quad &. \end{aligned}$$

(b) Proof by induction. Pick element n. If it forms a set by itself the remaining $n-1$ elements are to form $k-1$ partitions so that adding the set containing n we have k partitions for the n elements. If however element n is in a set of cardinality at least 2, the first generate a partition of the remaining $n-1$ elements into k sets and then add n into any of those k sets. Combining the two possibilities for n we have

$$S(n,k) = S(n-1,k-1) + k \cdot S(n-1,k).$$

Revisiting part (a). We can confirm that our enumeration is correct as follows.

$$S(3,2) = S(2,1) + 2S(2,2) = 1 + 2S(2,2) = 1 + 2 = 3.$$

■

Exercise 362.
Show for any $n \geq 2$ we have the following

$$S(n,2) = 2^{n-1} - 1.$$

Solution.
Base case for $n = 2$. $S(2,2) = 2^1 - 1 = 1$ true as noted before.
Inductive step. Let it be true for $n-1$ that is

$$S(n-1,2) = 2^{n-2} - 1.$$

Then

$$S(n,2) = S(n-1,1) + 2 \cdot S(n-1,2) = 1 + 2 \cdot (2^{n-2} - 1) = 2^{n-1} - 1,$$

and the result (inductive step) has been proven. ■

Exercise 363.
Show the following identities by combinatorial interpretation:

$$S(n+1,m+1) = \sum_{k=m}^{n} \binom{n}{k} S(k,m), \tag{5.5}$$

where $S(\cdot,\cdot)$ represents the Stirling number of the second kind.

Solution.
The Stirling number $S(n+1,m+1)$ counts the number of ways that $n+1$ distinct objects may be placed in $m+1$ indistinct bins, with at least one object per bin. Consider the bin in which object 1 is placed. There must be at least m objects not placed in this bin (since the other bins cannot be empty), and at most n (since that's how may other objects there are). Once we know how many objects are not placed in the same bin as 1, we may reduce the quantity to the number of ways we may chose those (k) objects, times the number of ways we may place those remaining objects in the remaining m bins with at least one object per bin. This is what is counted on the right hand side. ■

5.18 Bell numbers

Exercise 364.

Bell numbers. Let

$$B(n) = \sum_{k=1}^{n} S(n,k).$$

where $B(0) = 0$. Thus $B(n)$ is the total number of partitions of an n-element set. Show then the following for all $n \geq 1$.

$$B(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} B(k)$$

Solution.

Let $\{1,2,\ldots,n\}$ be the elements of the n-element set. Element n will appear in any of the n set of the partition along with k additional elements, where $k \geq 0$. The are $\binom{n-1}{k}$ ways to pick those additional elements; the remaining $n-1-k$ ways can be picked in $B(n-1-k)$ ways. Thus

$$B(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} B(n-k-1) = \sum_{j=0}^{n-1} \binom{n-1}{n-j-1} B(j) = \sum_{j=0}^{n-1} \binom{n-1}{j} B(j) = \sum_{k=0}^{n-1} \binom{n-1}{k} B(k).$$

■

Chapter 6

Recurrences and generating functions

6.1 Short review on recurrences

6.1.1 Arithmetic and geometric progressions

Definition 6.1 **Arithmetic Progression**

Let $a_0 = b$ and let $a_n = a_{n-1} + s$ otherwise. The initial value b is known as the base case value or boundary value. The increment is know as the step s. The sequence generated is

$$b, b+s, b+2s, b+3s, \ldots$$

Thus $a_n = b + ns$ for $n \geq 0$, where $a_0 = b$ and $a_n = a_{n-1} + s$ for $n > 0$.

Definition 6.2 **Geometric progression**

Let $a_0 = b$ and let $a_n = sa_{n-1}$ otherwise. The initial value b is known as the base case value or boundary value. The s is know as the growth factor (multiplicative step). The sequence generated is

$$b, b*s, b*s^2, b*s^3, \ldots$$

Thus $a_n = bs^n$ for $n \geq 0$,

6.1.2 Recurrence relations

A recurrence or recurrence relation describes how one can obtain the n-th term of a sequence from prior terms of the sequence. Let a sequence has its first term a 1. The next term is obtained by doubling the previous term. Thus if $a_0 = 1$, the $a_1 = 2 \cdot a_0$, $a_2 = 2 \cdot a_1$ and in general $a_{n+1} = 2a_n$ for $n \geq 0$. Equivalently $a_n = 2a_{n-1}$ for $n > 0$ i.e. $n \geq 1$. We can prove by induction that $a_n = 2^n$ (and in this latter case $n \geq 0$).

6.1.3 Recurrences of order k

Definition 6.3 **Recurrences of order k**

A recurrence relation of order k is a function of the form

$$a_n = f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$$

6.1.4 Linear Recurrences of order k

Definition 6.4 **Linear Recurrences of order k**

A linear recurrence relation of order k is a function of the form

$$a_n = f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$$

where $f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$ is a linear function i.e.

$$f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n) = A_{n-1}a_{n-1} + A_{n-2}a_{n-2} + \ldots + A_{n-k}a_{n-k} + g(n)$$

where $g(n)$ is a function of n, and $A_{n-1}, \ldots, A_{n-k}$ are constant coefficients. If $f(n) = 0$ it is called homegeneous.

We can easily compute (calculate) a_n if all of $a_{n-1}, a_{n-2}, \ldots, a_{n-k}$ known and if available, $f(n)$ for a given n. Thus if we are given a sequence of k initial values, for the first k elements of the sequence the sequence can be computed.

6.1.5 Ordinary generating functions

Definition 6.5 **Ordinary generating functions (o.g.f.)**

Let $a_0, \ldots, a_n, \ldots$ be a sequence of real numbers. Then $A(x)$ defined as follows is known as the ordinary generating function (o.g.f.) of sequence $a_n (n \geq 0)$.

$$A(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$$

We shall refer to $A(x)$ as the generating function of a_n.

6.1.6 Exponential generating functions

Definition 6.6 **Exponential generating functions (e.g.f.)**

Let $a_0, \ldots, a_n, \ldots$ be a sequence of real numbers Then $A(x)$ defined as follows is known as the exponential generating function (e.g.f.) of sequence $a_n (n \geq 0)$.

$$A(x) = a_0 + a_1 x + a_2 \frac{x^2}{2!} + \ldots + a_n \frac{x^n}{n!} + \ldots$$

6.2 Aritmetic progressions

Exercise 365.

Forward chain method. Let $a_n = a_{n-1} + s$ for $n > 0$ and $a_0 = b$. Solve this recurrence using the forward chain method.

Solution.

$$\begin{aligned} a_0 &= b \\ a_1 &= a_0 + s \\ a_2 &= a_1 + s \\ &\dots \\ a_{n-1} &= a_{n-2} + s \\ a_n &= a_{n-1} + s \end{aligned}$$

We add up all these equations. The number of equations is $n+1$. The left hand side of equation (0) has an a_0 that cancels with the right-hand side a_0 of equation (1). Likewise for $a_1, a_2, \dots, a_{n-1}$. Thus after all those cancellation on the left-hand side we are left with a_n and the right hand side we are left with a b and several s from equation (1) through equation (n), a total of n of them. Thus

$$a_n = b + n * s.$$

■

Exercise 366.

Backtracking chain method. Let $a_n = a_{n-1} + s$ for $n > 0$ and $a_0 = b$. Solve this recurrence using the backtracking chain known as backward substitution or the iteration method or the repeated substitution methods.

Solution.
To compute a_n, a_{n-1}, a_{n-2} i.e. to generate the right hand-side of each one of these terms we substitute in the reccurence $n, n-1, n-2$ for n. Then we generate the following terms

$$\begin{aligned} a_n &= a_{n-1} + s \\ a_{n-1} &= a_{n-2} + s \\ a_{n-2} &= a_{n-3} + s \\ &\dots \end{aligned}$$

We will be using these terms in the solution derived below.

$$\begin{aligned} a_n &= a_{n-1} + s \\ &= (a_{n-2} + s) + s \\ &= a_{n-2} + 2 \times s \\ &= (a_{n-3} + s) + 2 \times s \\ &= a_{n-3} + 3 \times s \\ &\dots \\ &= a_0 + n \times s \\ &= b + n \times s \end{aligned}$$

We unfold the recurrence. To compute a_n, a_{n-1}, a_{n-2} i.e. to generate the right hand-side of each of these terms we subsitute in the reccurence $n, n-1, n-2$ for n. We repeat until we encounter on the right-hand side the base case a_0 which is replaced by the base case value b obtained through the base case $a_0 = b$.

$$a_n = b + n * s.$$

■

6.3 Geometric progressions

Exercise 367.

Geometric progression. Let $a_0 = b$ and let $a_n = sa_{n-1}$ otherwise. The initial value b is known as the base case value or boundary value. The s is know as the growth factor (multiplicative step). The sequence generated is

$$b, b*s, b*s^2, b*s^3, \ldots$$

Thus $a_n = bs^n$ for $n \geq 0$, Solve the following recurrence with a forward chain like method or a backward chain method: The recurrence is $a_n = sa_{n-1}$ for $n > 0$ and $a_0 = b$, as stated earlier.

Solution.
We can solve this recurrence using the backward chain method.

$$\begin{aligned} a_n &= sa_{n-1} \\ a_{n-1} &= sa_{n-2} \\ &\ldots \\ a_3 &= sa_2 \\ a_2 &= sa_1 \\ a_1 &= sa_0 \end{aligned}$$

We multiply all equations. The topmost equation (1) has a a_{n-1} in its right-hand side that will cancel out with the left-hand side a_{n-1} of Equation (2). At the end we are left again with a_n on the left-hand side and n s's and of course a_0 which is b. Thus

$$a_n = s^n b = bs^n.$$

Note that implicitly we assumed that all cancelled out terms were not 0! ■

6.4 Generating functions

Exercise 368.

Generating functions can be manipulated as formal power series. Properties of Generating Functions.
Let $a_0, \ldots, a_n, \ldots$ be a sequence of real numbers and let $A(x)$ be its ordinary generating function. Show that $B(x) = xA(x)$ is the generating function of sequence $0, a_0, a_1, \ldots$ where $b_n = a_{n-1}$ for $n > 0$ and $b_0 = 0$.

Solution.

$$\begin{aligned} A(x) &= a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} + a_n x^n + \ldots \\ xA(x) &= 0 + a_0 x + a_1 x^2 + \ldots + a_{n-1} x^n + a_n x^{n+1} + \ldots \\ B(x) &= b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n + \ldots \end{aligned}$$

Thus $b_0 = 0, b_1 = a_0, b_2 = a_1, \ldots b_n = a_{n-1}$, as needed. ■

Exercise 369.

Let $a_0, \ldots, a_n, \ldots$ be a sequence of real numbers and let $A(x)$ be its ordinary generating function. Show that $B(x) = A(x) - a_0$ is the generating function of sequence $0, a_1, a_2, \ldots$ where $b_n = a_n$ for $n > 0$ and $b_0 = 0$.

Solution.

$$\begin{aligned} A(x) &= a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots \\ A(x) - a_0 &= 0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots \\ B(x) &= b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n + \ldots \end{aligned}$$

Thus $b_0 = 0, b_1 = a_1, b_2 = a_2, \ldots b_n = a_n$, as needed. ■

Exercise 370.

Let $a_0,\ldots,a_n,\ldots$ be a sequence of real numbers and let $A(x)$ be its ordinary generating function. Show that $B(x)=(A(x)-a_0)/x$ is the generating function of sequence $a_1,a_2,\ldots$ where $b_n=a_{n+1}$ for $n\geq 0$.

Solution.

$$\begin{aligned} A(x) &= a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots \\ (A(x)-a_0)/x &= a_1+a_2x+a_3x^2+\ldots+a_{n+1}x^n+\ldots \\ B(x) &= b_0+b_1x+b_2x^2+\ldots+b_nx^n+\ldots \end{aligned}$$

Thus $b_0=a_1,b_1=a_2,b_2=a_3,\ldots b_n=a_{n+1}$, as needed. ■

Exercise 371.
(a) Addition of two o.g.f. Let $a_0,\ldots,a_n,\ldots$ be a sequence of real numbers and let $A(x)$ be its ordinary generating function. Let $b_0,\ldots,b_n,\ldots$ be a sequence of real numbers and let $B(x)$ be its ordinary generating function. Then sequence $c_n=(a_n+b_n)$ for $n\geq 0$ has o.g.f $C(x)=A(x)+B(x)$.
(b) Multiplication of two of o.g.f. Let $a_0,\ldots,a_n,\ldots$ be a sequence of real numbers and let $A(x)$ be its ordinary generating function. Let $b_0,\ldots,b_n,\ldots$ be a sequence of real numbers and let $B(x)$ be its ordinary generating function. Then sequence $C(x)=A(x)B(x)$ is the generating function of the sequence

$$c_n=\sum_{i=0}^{n}a_ib_{n-i}$$

Counting problems can be restated as problems of determining the coefficient of x^i in a certain o.g.f.

Solution.
(a) Easy.
(b) Easy. ■

Exercise 372.
If $A(x)$ is the o.g.f for counting objects from set A, and if $B(x)$ is the o.g.f for counting objects from set B, then show $A(x)B(x)$ is the o.g.f. for counting objects from set $A \cup B$.

Solution.
If $A(x) = a_0 + a_1x + \ldots + a_ix^i$, and $B(x) = b_0 + b_1x + \ldots + b_ix^i$, then

$$A(x)B(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \ldots + \left(\sum_{k=0}^{i} f_k g_{i-k}\right) x^i + \ldots$$

■

Exercise 373.
If $A(x)$ is the e.g.f for counting objects from set A, and If $B(x)$ is the e.g.f for counting objects from set B, then show that $a_i b_j$ is the number of arrangements of two objects respectively, and the cofficient of x^m in $A(x)B(x)$.

Solution.
If $A(x) = \sum_i a_i x^i/i!$ and $B(x) = \sum_i b_i x^i/i!$ $A(x)B(x)$ is

$$A(x)B(x) = \sum_{m=0}^{\infty} \left(\sum_{i+j=m} \frac{a_i}{i!} \frac{b_j}{j!} \right) x^m$$

The number of ways of mixing i objects that are of one type to j objects that are of another type is $\frac{(i+j)!}{i!j!}$. Thus the e.g.f for mixing the two types of objects has coefficient x^m equal to (note $m = i+j$)

$$\frac{1}{m!} \cdot \sum_{i+j=m} a_i b_j \frac{(i+j)!}{i!j!}$$

■

Exercise 374.
What is the o.g.f. with $a_i = \binom{n}{i}$?

Solution.
The generation function of $a_i = \binom{n}{i}$ is $A(x) = \sum_{i=0}^{n} \binom{n}{i} x^i = (1+x)^n$. Note that

$$(1+x)^n = (1+x)(1+x)\dots(1+x)$$

Each contribution to the coefficient of x^i corresponds to a choice of i of the $(1+x)$ factors by choosing the x from them and multiplying them. ■

Exercise 375.

In how many ways can we pick r balls from a collection that contains at least r red balls and at least r blue balls? (Balls of the same color are indistinct, order does not matter.)

Solution.
Choosing r balls from a collection A of r or more indistinct red balls can be done in only one way, thus $a_i = 1$. (Another way to state this is that we have a collection A of an infinite number of red balls). Then $A(x) = 1+x+x^2+\ldots$. Likewise choosing r balls from a collection of r or more indistinct blue balls can be done in only one way, thus $b_i = 1$. Then $B(x) = 1+x+x^2+\ldots$. Choosing r balls from a collection that contains red and blue balls can be done in $r+1$ ways by picking $0,1,2,\ldots r$ red balls with the balance being blue balls . Thus $c_i = i+1$. Therefore $C(x) = 1+2x+3x^2+\ldots+(i+1)x^i$. Notice that

$$A(x)B(x) = \left(\sum_{i=0}^{\infty} x^i\right)\cdot\left(\sum_{i=0}^{\infty} x^i\right) = \sum_{i=0}^{\infty}(i+1)x^i$$

For the red balls choices of set A of red balls we have $A(x)1+x+x^2+\ldots+ = 1/(1-x)$. Likewise $B(x) = 1/(1-x)$. The $C(x) = A(x)B(x) = 1/(1-x)^2 = (1-x)^{-2}$. By the binomial theorem

$$(1+x)^m = \sum_{i=0}^{\infty}\binom{m}{i}x^i$$

where

$$\binom{m}{i} = \frac{m(m-1)\ldots(m-i+1)}{i!}$$

Moreover we can show that

$$C(x) = 1+2x+3x^2+\ldots+(i+1)x^i = (1-x)^{-2}$$

This is because

$$\left(1+2x+3x^2+\ldots+(i+1)x^i+\ldots\right)\cdot(1-x)^2 = \left(1+2x+3x^2+\ldots+(i+1)x^i+\ldots\right)\cdot(1-x^2+2x)$$

and thus it is equal

$$1+\sum_{i=1}^{\infty}\left((i+1)x^i-2ix^i+(i-1)x^i\right) = 1$$

■

Exercise 376.
Let $A(x) = 1/(1-x)$ be an ordinary generating function. A sequence with such a generating function is the sequence $a_n = 1$ for all n. This is because

$$A(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots = 1 + x + x^2 + \ldots + x^n + \ldots = 1/(1-x).$$

Consider now the sequence $b_n = k^n$ for all n where k is some constant. What is the generating function $B(x)$?

Solution.

$$B(x) \quad = \quad b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n + \ldots = k^0 + k^1 x + \ldots + k^n x^n + \ldots = \frac{1}{1-kx}$$

■

Exercise 377.
(a) Find the sequence that has $A(x) = (1+x)^2$.
(b) Find the o.g.f of $0,1,1,\ldots$.
(c) Find the o.g.f of $5^0, 5^1, 5^2, \ldots, 5^n, \ldots$.
(d) Find the e.g.f of $5^0, 5^1, 5^2, \ldots, 5^n, \ldots$.

Solution.
(a) Since the o.g.f $A(x)$ is $A(x) = 1+2x+2x^2$, then the sequence is $1,2,2,0,\ldots$.
(b) Since the sequence is $0,1,1,\ldots$, we conclude that the o.g.f. is $1+x+x^2+\ldots = 1/(1-x)$. This is $A(x) = 0+1\cdot x+1\cdot x^2+\ldots+ = 1/(1-x)-1 = x/(1-x)$.
(c) The o.g.f. is $1/(1-5x)$ similarly to the previous case.
(d) The e.g.f of $5^0, 5^1, 5^2, \ldots, 5^n, \ldots$ is $A(x) = 5^0 + 5^1 x + 5^2 x^2/2! + \ldots + 5^n x^n/n! + \ldots = e^{5x}$. ■

Exercise 378.

Find a closed form expression for the ordinary generating function (o.g.f.) of the sequence $a_n = n^2$.

Solution.
There are many ways to solve this problem. From

$$1 + x + x^2 + \ldots + x^n + \ldots = \frac{1}{1-x},$$

if we take the derivative of each side we get

$$0 + 1 + 2x + 3x^2 + 4x^3 + \ldots + nx^{n-1} + \ldots = \frac{1}{(1-x)^2} \tag{6.1}$$

and if we take the derivative of each side once more we get

$$2 + 3 \cdot 2x^1 + 4 \cdot 3x^2 + \ldots + n \cdot (n-1)x^{n-2} + \ldots = \frac{2}{(1-x)^3}.$$

Now, we multiply both sides of the last expression with x^2 and we get

$$2x^2 + 3 \cdot 2x^3 + 4 \cdot 3x^4 + \ldots + n \cdot (n-1)x^n + \ldots = \frac{2x^2}{(1-x)^3}. \tag{6.2}$$

If we multiply Eq.(6.1) with x and add the result to Eq.(6.2) we get that

$$1x + 2^2x^2 + 3^2x^3 + \ldots + n^2x^n + \ldots = \frac{2x^2}{(1-x)^3} + \frac{x}{(1-x)^2} = \frac{x(1+x)}{(1-x)^3}.$$

■

Exercise 379.

Application of Fibonacci. Consider the recurrence $a_{n+2} = a_{n+1} + a_n$ where $a_1 = a_0 = 1$. Show that

$$\binom{n+1}{0} + \binom{n}{1} + \binom{n-1}{2} + \cdots = a_{n+1}$$

(the sum goes up to the last term where $n+1-k \geq k$).

Solution.
One way to solve this problem (another way is simultaneous induction on n,k and Pascal triangle's identity), is using the o.g.f. of the Fibonacci numbers. This is as follows.

$$A(x) = \frac{1}{1-x-x^2} = \sum_{i=0} (x+x^2)^i.$$

The coefficient of x^{n+1} in the power series above is

$$\binom{n+1}{0} + \binom{n}{1} + \binom{n-1}{2} + \cdots$$

which is the coefficient of x^{n+1} in the o.g.f of the Fibonacci numbers i.e. a_{n+1}. ■

Exercise 380.
How many ways can four distinct dice be rolled so that they sum to 10?

Solution.
For each die, the ordinary generating function is

$$x+x^2+x^3+x^4+x^5+x^6 \quad = \quad x(1+x+x^2+x^3+x^4+x^5) \quad = \quad \tfrac{x(1-x^6)}{1-x}$$

The answer is the coefficient of x^{10} in $\left(\frac{x(1-x^6)}{1-x}\right)^4$, which is the same as the coefficient of x^6 in $\left(\frac{1-x^6}{1-x}\right)^4$. This generating function is equal to

$$(1-4x^6+6x^{12}-4x^{18}+x^{24})(1-x)^{-4}$$

The coefficient of x^6 in this is the coefficient of x^6 in $(1-x)^{-4}$, minus four times the constant term of $(1-x)^{-4}$. The answer is therefore

$$\tbinom{4+6-1}{6}-4\tbinom{4+0-1}{0} \quad = \quad \tbinom{9}{6}-4\tbinom{3}{0} \quad = \quad 80$$

■

Exercise 381.
Use generating functions, to determine the value of the sum

$$\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\ldots+n\binom{n}{n}$$

Solution.

$$f(x) = 1+\binom{n}{1}x+\binom{n}{2}x^2+\binom{n}{3}x^3+\ldots+\binom{n}{n}x^n$$

observe that

$$f'(x) = \binom{n}{1}+2\binom{n}{2}x+3\binom{n}{3}x^2+\ldots+n\binom{n}{n}x^{n-1}$$

The answer is therefore $f'(1)$. But $f(x)$ is simply $(1+x)^n$, so $f'(x)$ is $n(1+x)^{n-1}$. Thus the answer is $n2^{n-1}$. ■

Exercise 382.

Find a closed form expression for the ordinary generating function $\sum_{n=0}^{\infty} \alpha_n x^n$, where

$$\alpha_n = \sum_{k \geq 0} \binom{k}{n-k}$$

(Assume $\binom{r}{s}$ is 0 if $s > r$ or $s < 0$).

Solution.
If $f(x)$ is the generating function $\sum_{n=0}^{\infty} \alpha_n x^n$, then

$$\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} \sum_{k \geq 0} \binom{k}{n-k} x^n \\
&= \sum_{k \geq 0} \sum_{n=0}^{\infty} \binom{k}{n-k} x^n \\
&= \sum_{k \geq 0} x^k \sum_{n=0}^{\infty} \binom{k}{n-k} x^{n-k} \\
&= \sum_{k \geq 0} x^k \sum_{n=k}^{2k} \binom{k}{n-k} x^{n-k} \\
&= \sum_{k \geq 0} x^k (1+x)^k \\
&= \sum_{k \geq 0} (x+x^2)^k \\
&= \frac{1}{1-x-x^2}
\end{aligned}$$

■

Exercise 383.
We have a set A of 3 red balls and a set B of 4 green balls . In how many way can we choose two balls ?

Solution.
The generating function for $A(x)$ is $A(x) = 1+x+x^2+x^3$ and for $B(x) = 1+x+x^2+x^3+x^4$. Multiplying them we can read off the coefficient of x^2. ■

Exercise 384.
Find the coefficient of x^n in

$$\frac{x}{(1-x)(1-2x)^2}.$$

Solution.
We expand

$$\frac{x}{(1-x)(1-2x)^2}$$

as follows:

$$\frac{x}{(1-x)(1-2x)^2} = \frac{1}{(1-x)} + \frac{-1}{(1-2x)} + \frac{2x}{(1-2x)^2}$$

Therefore the coefficient of x^n in the original expression is the coefficient of x^n in $\frac{1}{1-x}$ which is 1 plus the coefficient of x^n in $\frac{-1}{1-2x}$ which is -2^n plus the coefficient of x^n in $\frac{2x}{(1-2x)^2}$ which is $n2^n$ (since the sequence $n = 1, 2, \ldots$ has o.g.f. $\frac{x}{(1-x)^2}$) Therefore the coefficient of x^n is $1 + (n-1)2^n$. ■

Exercise 385.
How many k letter words can be formed from the letter of the word MATHEMATICS?

Solution.
There are 2 M,A,T in MATHEMATICS and the 5 other characters appear once. Form $A(x) = (1+x+x^2)^3(1+x)^5$ and compute the coefficient of x^k. ■

Exercise 386.
In how many ways can we throw 15 indistint objects into 5 bins so that every bin gets at least 2 objects?

Solution.
Formulate $A(x) = (x^2 + x^3 + x^4 + \ldots)^5$ and calculate the coefficient of x^k. ■

Exercise 387.
How many ways can three distinct dice be thrown such that their sum is 9?

Solution.
The o.g.f for the outcome of a dice is $x+x^2+\ldots+x^6$. So for the three dice we have $(x+x^2+\ldots+x^6)^3$. We can write the above formula as

$$x^3(1+x+\ldots+x^5)^3 \ = x^3\frac{(1-x^6)^3}{(1-x)^3}$$

The coefficient of x^9 of the formula above or equivalently the coefficient of x^6 in $(1-x^6)^3\cdot(1-x)^{-3}$ is the answer to our problem. These two powers can be written as:

$$\begin{aligned}(1-x^6)^3 &= \sum_{i=0}^{3}\binom{n}{i}1^i(-1)^{3-i}x^{6(3-i)}\\ (1-x)^{-3} &= \sum_{r=0}^{\infty}\binom{r+3-1}{r}x^r.\end{aligned}$$

Therefore, the coefficient of the x^6 is equal to

$$\binom{3}{3}(-1)^{3-3}\binom{8}{6}+\binom{3}{2}(-1)^{3-2}\binom{2}{0}=28-3=25.$$

■

Exercise 388.

In how many ways can $3r$ balls be selected from $2r$ red balls, $2r$ blue balls, and $2r$ white balls?

Solution.
With generating functions. (A previous answer did not use them.) We must find the coefficient of x^{3r} in $(1+x+\ldots+x^{2r})^3$.

$$\begin{aligned}\left(1+x+\ldots+x^{2r}\right)^3 &= \left(\frac{1-x^{2r+1}}{1-x}\right)^3 \\ &= \left(1-3x^{2r+1}+3x^{4r+2}-x^{6r+3}\right)\frac{1}{(1-x)^3}\end{aligned}$$

The coefficient of x^{3r} in this mess is just the coefficient of x^{3r} in $(1-x)^{-3}$ minus three times the coefficient of x^{r-1} in $(1-x)^{-3}$, which is

$$\begin{aligned}\binom{3r+2}{3r}-3\binom{r+1}{r-1} &= \frac{(3r+2)(3r+1)}{2}-3\frac{(r+1)(r)}{2} \\ &= \frac{(9r^2+9r+2)-(3r^2+3r)}{2} \\ &= 3r^2+3r+1\end{aligned}$$

■

Exercise 389.

Two professors in two different subjects are giving oral examinations to 12 students at the same exam time. Each student will be examined individually for 5 minutes in each subject. In how many ways can a schedule be made up so that no student will have to see both professors at the same time? Give an approximate value to the answer as well.

Solution.
Suppose professor A examines students 1 through 12 in order. Then professor B may examine the students in any order that is a permutation of their indices where $\pi(i) \neq i$. This is just the number of derangements. Since A may examine the students in any order, the answer is

$$(12!) \cdot (12!) \cdot \sum_{l=0}^{12} \frac{(-1)^l}{l!} \quad \approx \quad \frac{(12!)^2}{e}$$

■

Exercise 390.

In how many ways can one select n letters from an unlimited supply of A's, B's, and C's so that no letter appears exactly three times?

Solution.
The generating function for each letter is

$$1+x+x^2+x^4+x^5+\ldots=\frac{1}{1-x}-x^3$$

The answer is therefore the coefficient of x^n in

$$(\frac{1}{1-x}-x^3)^3=\frac{1}{(1-x)^3}-\frac{3x^3}{(1-x)^2}+\frac{3x^6}{1-x}-x^9$$

This is equal to the coefficient of x^n in $(1-x)^{-3}$ $(\binom{n+2}{n})$, minus three times the coefficient of x^{n-3} in $(1-x)^{-2}$ $(\binom{n-2}{n-3})$ if $n\geq 3$, plus three times the coefficient of x^{n-6} in $(1-x)^{-1}$ (one) if $n\geq 6$, minus one in the case where $n=9$. These combine to equal

$$\begin{cases} \frac{(n+2)(n+1)}{2}-3(n-2)+3=\frac{n^2-3n+20}{2} & \text{if } n\geq 6 \text{ and } n\neq 9 \\ \frac{(n+2)(n+1)}{2}-3(n-2)+2=36 & \text{if } n=9 \\ \frac{(n+2)(n+1)}{2}-3(n-2)=\frac{n^2-3n+14}{2} & \text{if } 5\geq n\geq 3 \\ \frac{(n+2)(n+1)}{2} & \text{if } 2\geq n\geq 0 \end{cases}$$

■

Exercise 391.

In how many ways can one select n letters from an unlimited supply of A's, B's, and C's so that there are an even number of A's (zero counts as an even number)?

Solution.

Using generating functions, the answer is the coefficient of x^n in $(1+x^2+x^4+\ldots)(1+x+x^2+\ldots)^2$, where the first term counts the number of ways of selecting the A's, and the second the B's and C's. This coefficient is equal to

$$\begin{aligned}
\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \text{coefficient of } n-2i \text{ in } \frac{1}{(1-x)^2} &= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-2i+1}{n-2i} \\
&= \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} n-2i+1 \\
&= (\lfloor \frac{n}{2} \rfloor + 1)(n+1) - 2\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} i \\
&= (\lfloor \frac{n}{2} \rfloor + 1)(n+1) - \lfloor \frac{n}{2} \rfloor (\lfloor \frac{n}{2} \rfloor + 1) \\
&= (\lfloor \frac{n}{2} \rfloor + 1)(n - \lfloor \frac{n}{2} \rfloor + 1) \\
&= \begin{cases} \left(\frac{n}{2}+1\right)^2 & \text{if } n \text{ is even} \\ \left(\frac{n+1}{2}\right)\left(\frac{n+3}{2}\right) & \text{if } n \text{ is odd} \end{cases}
\end{aligned}$$

■

Exercise 392.

Find a closed form expression for the ordinary generating function of the sequence $a_n = n^3$.

Solution.
We shall start of with the relation

$$f(x) = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots$$

for which we know $f(x) = (1-x)^{-1}$ and work from there.

$$\begin{array}{rcccl}
f'(x) & = & \frac{1}{(1-x)^2} & = & 1 + 2x + 3x^2 + \ldots + nx^{n-1} + \ldots \\
xf'(x) & = & \frac{x}{(1-x)^2} & = & x + 2x^2 + 3x^3 + \ldots + nx^n + \ldots \\
(xf'(x))'(x) & = & \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3} & = & 1 + 4x + 9x^2 + \ldots + n^2x^{n-1} + \ldots \\
x(xf'(x))'(x) & = & \frac{x}{(1-x)^2} + \frac{2x^2}{(1-x)^3} & = & x + 4x^2 + 9x^3 + \ldots + n^2x^n + \ldots \\
(x(xf'(x))')'(x) & = & \frac{1}{(1-x)^2} + \frac{6x}{(1-x)^3} + \frac{6x^2}{(1-x)^4} & = & 1 + 8x + 27x^2 + \ldots + n^3x^{n-1} + \ldots \\
x(x(xf'(x))')'(x) & = & \frac{x}{(1-x)^2} + \frac{6x^2}{(1-x)^3} + \frac{6x^3}{(1-x)^4} & = & x + 8x^2 + 27x^3 + \ldots + n^3x^n + \ldots
\end{array}$$

That last series is what we want, so the answer is $\frac{x}{(1-x)^2} + \frac{6x^2}{(1-x)^3} = \frac{x^3+4x^2+x}{(1-x)^4}$. ■

Exercise 393.

In how many ways can one form n digit words (i.e. order matters) from the set $\{0,1,2,3\}$ in which the number of 0's is even?

Solution.
For this problem we want to consider exponential generating functions. For $\{1,2,3\}$ the e.g.f.'s are

$$1+x+\frac{x^2}{2}+\frac{x^3}{6}+\ldots+\frac{x^n}{n!}+\ldots=e^x$$

For 0, the e.g.f. is

$$1+\frac{x^2}{2}+\frac{x^4}{24}+\ldots+\frac{x^{2n}}{(2n)!}+\ldots=\frac{e^x+e^{-x}}{2}$$

That is true because the expansions of e^x and e^{-x} both include $\frac{x^i}{i!}$ for i even, but for i odd the latter has $-\frac{x^i}{i!}$ and the two terms cancel out. Thus the answer is the coefficient of $\frac{x^n}{n!}$ in

$$(e^x)^3\frac{e^x+e^{-x}}{2}=\frac{e^{4x}+e^{2x}}{2}$$

This coefficient is $\frac{4^n+2^n}{2}=2^{2n-1}+2^{n-1}$. ■

Exercise 394.

Partitions revisited. We have indistinct objects into indistinct bins. The number of partitions of n is the coefficient of x^n in the following o.g.f $A(x)$. $A(x)$ is the product of o.g.f. of the form

$$\frac{1}{1-x^t} = 1 + x^{1 \cdot t} + x^{2 \cdot t} + \ldots + x^{i \cdot t} + \ldots$$

The term $x^{i \cdot t}$ contributes once i instances of t in the partition. Thus

$$A(x) = (1+x+x^2+x^3+\ldots)(1+x^2+x^4+x^6+\ldots)(1+x^3+x^6+x^9+\ldots)(1+x^4+x^8+x^12+\ldots)(1+x^5+x^10+x^15+\ldots)$$

Find the partitions of 5.

Solution.
From the o.g.f $A(x)$ above the partitions of 5 and the number of partitions of 5 are determined during the computation of the coefficient of x^5 in $A(x)$.
We observe that an x^5 can be generated as follows:
(1) x^5 from fifth term. This is 5.
(2) x^4 from fourth term and x^1 from first term. This is $4+1 \cdot 1 = 4+1$.
(3) x^3 from third term and x^2 from second term. This is $3+2 \cdot 1 = 3+2$.
(4) x^3 from third term and x^2 from first term. This is $3+1 \cdot 2 = 3+1+1$.
(5) x^4 from second term and x^1 from first term. This is $2 \cdot 2+1 \cdot 1 = 2+2+1$.
(6) x^2 from second term and x^3 from first term. This is $2 \cdot 1+1 \cdot 3 = 2+1+1+1$.
(7) x^5 from first term. This is $1 \cdot 5 = 1+1+1+1+1$. ■

Exercise 395.
210 chairs are to be arranged in 4 classrooms. Each classroom can get $0, 30, 60, 90$ chairs. Rooms are distinguishable. In how many ways can we distribute the chairs?

Solution.
Find the coefficient of x^{210} in

$$(1+x^{30}+x^{60}+x^{90})^4$$

■

Exercise 396.
A country has 9cent, 17cent, 31cent, and 100cent coins. In how many different ways can the 100cent coin be changed?

Solution.
Find the coefficient of x^{100} in

$$\frac{1}{1-x^9}\frac{1}{1-x^{17}}\frac{1}{1-x^{31}}$$

There is at least one way!

$$100 = 1\cdot 31 + 3\cdot 17 + 2\cdot 9$$

■

Exercise 397.
What is the number of n digit sequences consisting of $0,1,2,3$ with least one $2,3$?

Solution.
The e.g.f of 0 and 1 is $\sum_i x^i/i! = e^x$. The e.g.f of 2 and 3 is likewise $e^x - 1$. Thus $(e^x-1)^2 e^{2x}$ has coefficient x^n which is the coefficient of $x^n/n!$ in

$$(e^x-1)^2 e^{2x} = (e^{2x} - 2e^x + 1)e^{2x} = e^{4x} - 2e^{3x} + e^{2x}$$

which is

$$4^n - 23^n + 2^n.$$

■

Exercise 398.
In how many ways can we distribute m distinct objects into n distinct bins so that there is at least one object in each bin. Order in a bin does not matter.

Solution.
Each bin has e.g.f. $e^x - 1$ and collectively $(e^x - 1)^n$. Rest is left as an exercise. ■

Exercise 399.

How many ways can a convex n-gon be divided into triangles by non-intersecting diagonals ? First find the recurrence relation for the problem and solve it.

Solution.
Let a_n be the number of ways to divide up an n-gon. In any division, vertex 1 and vertex n form a triangle with vertex i for some $2 \leq i \leq n-1$. For each i we can recursively determine the number of such divisions, and the possibilities for each i are disjoint. Thus

$$a_n = \sum_{i=2}^{n-1} a_i a_{n-i+1} \ , \quad a_2 = 1$$

Let $A(x)$ be the generating function for this sequence.

$$\begin{aligned} \sum_{n=3}^{\infty} a_n x^n &= \sum_{n=3}^{\infty} \sum_{i=2}^{n-1} a_i a_{n-i+1} x^n \\ A(x) - a_2 x^2 &= \sum_{n=3}^{\infty} \sum_{i=2}^{n-1} a_i a_{n-i+1} x^n \end{aligned}$$

Note that

$$\begin{aligned} \frac{A^2(x)}{x} &= (a_2 x^2 + a_3 x^3 + \cdots)(a_2 x + a_3 x^2 + \cdots) \\ &= \sum_{n=3}^{\infty} x^n \sum_{i=2}^{n-1} a_i a_{n-i+1} \end{aligned}$$

Thus

$$\begin{aligned} A(x) - x^2 &= \frac{A^2(x)}{x} \\ A^2(x) - xA(x) + x^3 &= 0 \\ A(x) &= \frac{x \pm \sqrt{x^2 - 4x^3}}{2} \\ &= x \, \frac{1 \pm \sqrt{1 - 4x}}{2} \end{aligned}$$

So the answer is the coefficient of x^{n-1} in the fraction. This is

$$a_n = \frac{1}{n-1} \binom{2n-4}{n-2}$$

■

Exercise 400.
Say the patrons of a coffeeshop order k types of coffee. How many ways n patrons can do so ?
Simpler question: There are 6 patrons and 3 types of coffee: filtered, espresso, instant.

Solution.
Using generating functions for filtered coffee we have

$$1+x^1+x^2+\dots$$

Likewise for espresso,

$$1+y^1+y^2+\dots$$

and instant

$$1+z^1+z^2+\dots$$

The coefficients is the number of orders of the corresponding type. We multiply them together: the coefficient of $x^i y^j z^k$ indicates how many orders include i for filtered, j for espresso and k for instant coffee. The number of patrons is $i+j+k$ and this is supposed to be n. An alternative convenient way is set $x=y=z$ and again multiply the three generating function, but now asks for the coefficient of x^n. ■

Exercise 401.
What is the number of non-negative integer solutions of

$$x_1 + x_2 + x_3 = 10$$

with $x_1 \geq 5, x_2 \geq 1, 0 \leq x_3 \leq 2$?

Solution.
For x_1 we have $5, \ldots, 10$, i.e. its generating function is

$$x^5 + \ldots + x^{10}.$$

For x_2, and x_3 accordingly. Multiply the three together and find the coefficient of x^{10} in the product. ■

Exercise 402.

By a *pile* of coins we mean an arrangement of n coins in rows such that (see also attached figure):
(i) the coins in the first row form a single contiguous block,
(ii) in each higher row each coin touches exactly two coins from the row beneath it,
(iii) and *every* row consists of a single contiguous block of coins.
Let f_k be the number of piles that have a first row consisting of exactly k coins.
(a) Show that:

$$f_k = \sum_{j=1}^{k-1} (k-j) f_j \ + \ 1, \ (k = 1,2,\ldots) \text{ and } \ f_0 = 1.$$

Solution.
(a) If the first row consists of k coins, we can place on top of it any number of j coins, where $1 \leq j \leq k-1$. We can place j coins in $j-k$ ways on top of the first row (because there are so many slots the leftmost coin of a contiguous block of j coins can occupy). As soon as we place j coins in row 2, we can place on top of this row coins in f_j different ways (inductively). If $j = 0$ (no coins are placed) we get one configuration, which consists of one row only. Therefore, we get the recurrence (we take $f_0 = 1$ as a convention to denote the empty arrangement):

$$f_k = \sum_{j=1}^{k-1} (k-j) f_j + 1 = \sum_{j=1}^{k} (k-j) f_j + 1 = (\sum_{j=0}^{k} (k-j) f_j) - k f_0 + 1 = (\sum_{j=0}^{k} (k-j) f_j) - k + 1$$

(b) We will now find the o.g.f of f_j. We multiply both sides of it by x^k and sum for all values of k from 1 to ∞:

$$\sum_{k=1}^{\infty} f_k x^k = \sum_{k=1}^{\infty} ((\sum_{j=0}^{k} (k-j) f_j) - k f_0 + 1) x^k \Rightarrow F(x) - 1 = \sum_{k=1}^{\infty} \sum_{j=0}^{k} (k-j) f_j - \sum_{k=1}^{\infty} k x^k + \sum_{k=1}^{\infty} x^k$$

where, $F(x) = \sum_{i=0}^{\infty} f_i x^i$ and we know that $\sum_{j=0}^{\infty} j x^j = x/(1-x)^2$, and $f_0 = 1$. The double sum is just the product of the o.g.f. of the sequence $j, j \geq 0$ and the sequence $f_j, j \geq 0$ (from the multiplication rule of o.g.f.) and thus:

$$F(x) = (x/(1-x)^2) F(x) - x/(1-x)^2 + 1/(1-x) \Rightarrow F(x) = \frac{1-2x}{x^2 - 3x + 1}$$

■

Exercise 403.

There is a round table to seat n couples (call them husband and wife) so that no husband and wife seat next to each other. In how many ways can this be accomplished?

Solution.
Number the $2n$ seats clock-wise. Husbands occupy all odd-indexed or all even-indexed seats. There are $n!$ permutation of them getting seated at odd-indexed seats. Likewise for even-indexed seats for a total of $2 \cdot n!$. For each such arrangement there is going to be $f(n)$ ways to have the wifes seated. If a husband is seated at k, his wife cannot be seated at $k+1$ or $k-1$ (in circular order). Then the number of permutation $G(n)$ for the n pairs of husband and wife pair placement is

$$G(n) = n! - A(1)(n-1)! + A(2)(n-2)! - \ldots + (-1)^n A(n)0!$$

In this expression $A(i)$ indicates the property that i pairs are seated next to each other. $A(k)$ is the number of circular arrangements so that no two of the k chosen items (or numbers mapped to seating positions are adjacent i.e. consecutive). From a prior problem this is $\binom{2n-k+1}{k}$ if we have no circular arrangement (it is a flat arrangement). If we do have then from the number above we need to exclude $1, 2n$ that are consecutive in a circular arrangement of $1, 2, \ldots 2n$. Then $2, 2n-1$ cannot be chosen. Thus the $k-2$ non consecutive are to be drawn from $3, \ldots, 2n-2$ a total of $2n-4$ elements. There are then $\binom{2n-4-(k-2)+1}{k-2} = \binom{2n-k-1}{k-2}$ choices. Thus

$$A(k) = \binom{2n-k+1}{k} - \binom{2n-k-1}{k-2} = \frac{2n}{2n-k}\binom{2n-k}{k}$$

■

Exercise 404.

In how many ways can one select a set of $2r$ balls from an urn that contains r identical white balls, r identical blue balls, $2r$ identical green balls, and $3r$ identical red balls? (The solution should be a closed form expression in r, but don't worry about purely algebraic simplification).

Solution.
The generating function for the white and blue balls is $1+x+\ldots+x^r = \frac{1-x^{r+1}}{1-x}$. For the green balls it is $\frac{1-x^{2r+1}}{1-x}$, and for the red balls it is $\frac{1-x^{3r+1}}{1-x}$. Thus the answer is the coefficient of x^{2r} in

$$\frac{(1-x^{r+1})(1-x^{r+1})(1-x^{2r+1})(1-x^{3r+1})}{(1-x)^4}$$

We can ignore the effect of higher order terms in the numerator, so this is the same as the coefficient of x^{2r} in

$$\frac{1-2x^{r+1}}{(1-x)^4}$$

This is the coefficient of x^{2r} in $(1-x)^{-4}$, minus twice the coefficient of x^{r-1} in $(1-x)^{-4}$. This equals

$$\binom{2r+3}{3} - 2\binom{r+2}{3}$$

■

Exercise 405.
Using generating functions (and not induction) prove the following equality for all $n \geq 0$, where f_n is the n'th Fibonacci number:

$$f_0 + f_1 + f_2 + \ldots + f_n = f_{n+2} - 1$$

(Recall that f_n is the coefficient of x^n in the generating function $F(x) = \frac{1}{1-x-x^2}$).

Solution.
Multiplying both sides by x^n and summing for $n = 0$ to ∞, for the right hand side we get

$$\begin{aligned}
\sum_{n=0}^{\infty} f_{n+2}x^n - \sum_{n=0}^{\infty} x^n &= \frac{F(x)-x-1}{x^2} - \frac{1}{1-x} \\
&= \frac{\frac{1}{1-x-x^2}-x-1}{x^2} - \frac{1}{1-x} \\
&= \frac{1-x+x^2+x^3-1+x+x^2}{x^2(1-x-x^2)} - \frac{1}{1-x} \\
&= \frac{2+x}{1-x-x^2} - \frac{1}{1-x} \\
&= \frac{(2+x)(1-x)-(1-x-x^2)}{(1-x-x^2)(1-x)} \\
&= \frac{F(x)}{1-x}
\end{aligned}$$

For the left hand side we get

$$\sum_{n=0}^{\infty} (f_0 + f_1 + \ldots + f_n)x^n = \frac{F(x)}{1-x}$$

■

6.5 Recurrences

Exercise 406.

Recurrences of order k. A recurrence relation of order k is a function of the form

$$a_n = f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$$

Linear Recurrences of order k. A linear recurrence relation of order k is a function of the form

$$a_n = f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$$

where $f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$ is a linear function i.e.

$$f(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n) = A_{n-1}a_{n-1} + A_{n-2}a_{n-2} + \ldots + A_{n-k}a_{n-k} + g(n)$$

where $g(n)$ is a function of n, and $A_{n-1}, \ldots, A_{n-k}$ are constant coefficients. If $f(n) = 0$ it is called homegeneous. We can easily compute (calculate) a_n if all of $a_{n-1}, a_{n-2}, \ldots, a_{n-k}$ known and if available, $f(n)$ for a given n. Thus if we are given a sequence of k initial values, for the first k elements of the sequence the sequence can be computed.

Solve a linear recurrence of order k as shown earlier in its general form using the characteristic polynomial method.

Solution.
Let

$$a_n = A_{n-1}a_{n-1} + A_{n-2}a_{n-2} + \ldots + A_{n-k}a_{n-k} \tag{6.3}$$

be an order k homogeneous linear recurrence where A_i are constant and $\neq 0$. In order to be able to find a solution, the values of $a_0, \ldots, a_{k-1}$ must be given and then we can compute a_n for all $n \geq k$. The characteristic polynomial (equation) is

$$\phi(x) = x^k - A_{n-1}x^{k-1} - \ldots - A_{n-k}x^0 == x^k - A_{n-1}x^{k-1} - \ldots - A_{n-k}.$$

The roots of $\phi(x)$ are the characteristic roots of the recurrence relation.
(a) If a is a solution of $\phi(x)$, then a^n is a solution of Equation 6.3.
(b) If a is a solution of $\phi(x)$ of multiplicity q, then $a^n, na^n, n^2a^n, \ldots, n^{q-1}a^n$ are solutions of Equation 6.3. ■

Exercise 407.

Linear Recurrence solution using o.g.f.

Let $a_n = La_{n-1} + Ma_{n-2}$, $L \neq 0, M \neq 0$. Find $A(x)$ in terms of a_0 and a_1 and then show how to obtain a_n leading thus to the solution of the recurrence.

Solution.

We write the equation as follows. We multiply both sides by $x^2, x^3, \ldots$, and then add all the terms. Note that $A(x) = \sum_{n=0}^{\infty} a_n x^n$.

$$\begin{aligned}
a_n - La_{n-1} - Ma_{n-2} &= 0 \\
a_n x^n - La_{n-1}x^n - Ma_{n-2}x^n &= 0 \\
\sum_{n=2}^{\infty} a_n x^n - \sum_{n=2}^{\infty} La_{n-1}x^n - \sum_{n=2}^{\infty} Ma_{n-2}x^n &= 0 \\
\sum_{n=2}^{\infty} a_n x^n - Lx\sum_{n=2}^{\infty} a_{n-1}x^{n-1} - Mx^2\sum_{n=2}^{\infty} a_{n-2}x^{n-2} &= 0 \\
(A(x) - a_0 - a_1 x) - Lx(A(x) - a_0) - Mx^2 A(x) &= 0 \\
(1 - Lx - Mx^2)A(x) - a_0 - a_1 x + Lxa_0 &= 0 \\
A(x) &= \frac{a_0 + a_1 x + La_0 x}{1 - Lx - Mx^2}.
\end{aligned}$$

From that point on we use properties of o.g.f. to find the a_n with $A(x)$ as given above.

■

Exercise 408.
(a) Let $a_n = La_{n-1} + Ma_{n-2}$, $L \neq 0, M \neq 0$. Solve it.
(b) Let $a_n = 3a_{n-1} - 2a_{n-2}$. Let $a_0 = 1$ and $a_1 = 2$. Solve it.
(c) $a_n = 4a_{n-1} - 4a_{n-2}$, with $a_0 = 1, a_1 = 2$. Solve it.

Solution.
(a) Then $a_n - La_{n-1} - Ma_{n-2} = 0$, gives $\phi(x) = x^2 - Lx - M$. Let the roots be distinct and r_1 and r_2. Then $a_n = A \cdot r_1^n + B \cdot r_2^n$, where A, B can be computed from other information e.g. a_0, a_1.
(b) Then $\phi(x) = x^2 - 3x + 2$. We have $\phi(x) = (x-1)(x-2)$. Thus $r_1 = 1, r_2 = 2$ and $a_n = A \cdot 1^n + B \cdot 2^n = A + B2^n$.
Since $a_0 = 1 = A + B2^0 = A + B$.
Since $a_1 = 2 = A + B2^1 = A + 2B$.
By subtracting we derive $B = 1$ and then $A = 0$. Thus $a_n = 2^n$. We can confirm that for $a_n = 2^n$, we have

$$a_n = 3a_{n-1} - 2a_{n-2} = 3 \cdot 2^{n-1} - 2 \cdot 2^{n-2} = 3 \cdot 2^{n-1} - 2^{n-1} = 2 \cdot 2^{n-1} = 2^n.$$

(c) Then $\phi(x) = x^2 - 4x + 4$. Thus the roots of $\phi(x)$ is 2 with multiplicity 2. Therefore $a_n = A \cdot 2^n + B \cdot n \cdot 2^n$.
Since $a_0 = 1$ we have $a_0 = 1 = A + B \cdot 0 \cdot 1 = A$. Since $a_1 = 2$ we have $a_1 = 2 = A \cdot 2^1 + B \cdot 1 \cdot 2^1 == 2A + 2B$ which implies $B = 0$. Thus $a_n = 2^n$. We confirm $a_n = 2^n = 4 \cdot 2^{n-1} - 4 \cdot 2^{n-2} = 2 \cdot 2^n - 2^n = 2^n$. ■

Exercise 409.

Fibonacci. Let $f_n = f_{n-1} + f_{n-2}$ for $n > 1$, with $f_0 = 0$ and $f_1 = 1$ for $0 \leq n \leq 1$ which can also be summarialy describes as $f_i = i$ for $0 \leq i \leq 1$. The Fibonacci sequence is a 2-nd order linear and homogeneus recurrence since $f(n) = 0$.

Solution.
The solution f_n for $f_n = f_{n-1} + f_{n-2}$, for $n > 1$, with $f_0 = 0$ and $f_1 = 1$ is given below.

$$f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

Let $F(x) = \sum_{i=0} f_n x^n$.

$$F(x) = f_0 + f_1 x + f_2 x^2 + \ldots f_n x^n + \ldots$$

The characteristic equation derived from the recurrence $f_n = f_{n-1} + f_{n-2}$ is $f_n - f_{n-1} - f_{n-2} = 0$ i.e. $\phi(x) = x^2 - x - 1$ Let r_1 and r_2 be the two roots of $\phi(x) = 0$. $r_1 = (1+\sqrt{5})/2$ and $r_2 = (1-\sqrt{5})/2$. Then

$$f_n = Ar_1^n + Br_2^n = A(r_1^n - r_2^n)$$

Give that $f_0 =$ and $f_1 = 1$ we have that $0 = A + B$ and $1 = Ar_1 + Br_2$. This mean $B = -A$ and $1 = Ar_1 - Ar_2$. Then $A = 1/\sqrt{5}$ and $B = -1/\sqrt{5}$. Further details are left out and can be worked out easily. ■

Exercise 410.
Let $a_n = 2a_{n-1}a_{n-2}$. Solve it. (Hint: How do we solve it ? It is not a linear recurrence.)

Solution.
If we take logarithms, $\lg a_n = \lg_{a_{n-1}} + \lg_{a_{n-2}} + 1$. Set $\lg a_n$ to A_n and we have $A_n = A_{n-1} + A_{n-2} + 1$. The latter can become if we add $+1$ to both sides $A_n + 1 = (A_{n-1} + 1) + (A_{n-2} + 1)$. If $F_n = A_n + 1$, we have $F_n = F_{n-1} + F_{n-2}$ a suspiciously familiar recurrence! ■

Exercise 411.

Fibonacci once more. Solve it using o.g.f. (Let for $n > 1$: $f_n = f_{n-1} + f_{n-2}$. We also have $f_0 = 0$ and $f_1 = 1$.)

Solution.
We write the equation as follows. We multiply both sides by $x^2, x^3, \ldots$, and then add all the terms. Note that $F(x) = \sum_{n=0}^{\infty} f_n x^n$. In the last equation we use $f_0 = 0$ and $f_1 = 1$.

$$\begin{aligned}
f_n - f_{n-1} - f_{n-2} &= 0 \\
f_n x^n - f_{n-1} x^n - f_{n-2} x^n &= 0 \\
\sum_{n=2}^{\infty} f_n x^n - \sum_{n=2}^{\infty} f_{n-1} x^n - \sum_{n=2}^{\infty} f_{n-2} x^n &= 0 \\
\sum_{n=2}^{\infty} f_n x^n - x \sum_{n=2}^{\infty} f_{n-1} x^{n-1} - x^2 \sum_{n=2}^{\infty} f_{n-2} x^{n-2} &= 0 \\
(A(x) - f_0 - f_1 x) - x(F(x) - f_0) - x^2 F(x) &= 0 \\
(1 - x - x^2) F(x) - f_0 - f_1 x + x f_0 &= 0 \\
F(x) &= \frac{f_0 + f_1 x + L f_0 x}{1 - Lx - Mx^2} \\
F(x) &= \frac{x}{1 - x - x^2}.
\end{aligned}$$

From that point on we use properties of o.g.f. to find the a_n with $F(x)$ as given above. Details appear in the next example. ■

Exercise 412.

Fibonacci continued! Instead of starting with the recurrence, we may start with the generating function $A(x)$ for the given recurrence since $A(x) = a_0 + a_1x + \ldots + a_nx^n = \sum n = 0^{\infty} a_n x^n$. Things can also grow out of hand if we are not careful with the index bounds of the sum.
The solution f_n for $f_n = f_{n-1} + f_{n-2}$, for $n > 1$, with $f_0 = 0$ and $f_1 = 1$ is given below.

Solution.
Let $F(x) = \sum_{i=0} f_n x^n$.

$$F(x) = f_0 + f_1x + f_2x^2 + \ldots f_nx^n + \ldots$$

$$\begin{aligned}
F(x) &= \sum_{n=0}^{n=\infty} f_n x^n \\
F(x) &= f_0 + f_1x + \sum_{n=2}^{n=\infty} f_n x^n \\
F(x) &= f_0 + f_1x + \sum_{n=2}^{n=\infty} (f_{n-1} + f_{n-2})x^n \\
F(x) &= 0 + x + \sum_{n=2}^{n=\infty} f_{n-1}x^n + \sum_{n=2}^{n=\infty} f_{n-2}x^n \\
F(x) &= 0 + x + x \cdot \sum_{n=2}^{n=\infty} f_{n-1}x^{n-1} + x^2 \cdot \sum_{n=2}^{n=\infty} f_{n-2}x^{n-2} \\
F(x) &= 0 + x + x \cdot \sum_{n=1}^{n=\infty} f_n x^n + x^2 \cdot \sum_{n=0}^{n=\infty} f_n x^n \\
F(x) &= 0 + x + x(F(x) - f_0) + x^2F(x) \\
F(x)(1 - x - x^2) &= x \\
F(x) &= \frac{x}{1 - x - x^2} \\
F(x) &= \frac{x}{(1 - g_1x)(1 - g_2x)}
\end{aligned}$$

The two solutions $1/g_1, 1/g_2$ of the quadratic equation $1 - x - x^2$ are such that $(1 - g_1x)(1 - g_2x) = 1 - (g_1 + g_2)x + g_1g_2x^2$ i.e. $g_1 + g_2 = 1$ and $g_1g_2 = -1$. Also $g_2 - g_1 = \sqrt{5}$. Note that

$$\begin{aligned}
F(x) &= \frac{x}{(1 - g_1x)(1 - g_2x)} \\
F(x) &= \frac{A}{1 - g_1x} + \frac{B}{1 - g_2x} \\
\sum_n f_n x^n &= \frac{1}{(g_1 - g_2)} \frac{1}{1 - g_1x} + \frac{1}{(g_2 - g_1)} \frac{1}{1 - g_2x} \\
\sum_n f_n x^n &= \frac{1}{(g_1 - g_2)} \sum_n g_1^n x^n + \frac{1}{(g_2 - g_1)} \sum_n g_2^n x^n \\
f_n &= \frac{1}{g_1 - g_2} g_1^n + \frac{1}{g_2 - g_1} g_2^n
\end{aligned}$$

The penultimate line is derived through the example Proposition 6.4 with $k = g_1$ and $k = g_2$ respectively. Finally

a bit of clean up is in order.

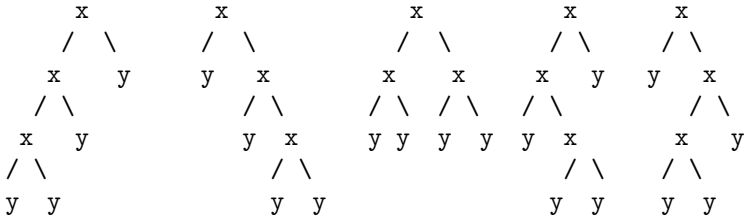
$$\begin{aligned} f_n &= \frac{1}{g_1 - g_2} g_1^n + \frac{1}{g_2 - g_1} g_2^n \\ f_n &= \frac{1}{g_2 - g_1} (g_2^n - g_1^n) \\ f_n &= \frac{1}{\sqrt{5}} (g_2^n - g_1^n) \end{aligned}$$

■

Exercise 413.

Catalan numbers revisited. Warm-up. Find the number of ways of pairwise parenthesizing 4 objects. Also count the number of full binary trees with 4 leaves. For example for $n = 4$ we have

```
(a((bc)d)), \\
(((ab)c)d), \\
(a(b(cd))), \\
((ab)(cd)).
```



Generalize and solve the problem for n objects or leaves!

Solution.
Let a_n be the number. Then the recurrence for a_n works as follows

$$a_n = a_{n-1}a_1 + a_{n-2}a_2 + \ldots + a_1a_{n-1} = \sum_{i=1}^{n-1} a_i a_{n-i}$$

noting that $a_0 = 0$ and $a_1 = 1$ Let $A(x) = \sum_{n=0}^{\infty} a_n x^n$.

$$\begin{aligned} A(x) - a_1x - a_0 &= \sum_{n=2}^{\infty} a_n x^n \\ \sum_{n=2}^{\infty} a_n x^n &= \sum_{n=2}^{\infty} \left(\sum_{i=1}^{n-1} a_i a_{n-i} \right) x^n \\ A(x) - a_1x - a_0 &= A^2(x) \end{aligned}$$

Thus $A^2(x) - A(x) + x = 0$ given $a_0 = 0$ and $a_1 = 1$. Solve for $A(x)$ i.e. $A(x) = \frac{1 \pm \sqrt{1-4x}}{2}$. Ignore the sign that generates negative a_n terms. Then by the binomial theorem we have

$$(1-4x)^{1/2} = \frac{1/2(1/2-1)\ldots(1/2-n+1)}{n!}(-4x)^n = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-3)}{n!} 2^n x^n$$

Then

$$\frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-3)}{n!} 2^n x^n = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-3)}{n!} \frac{1 \cdot 2 \cdot 3 \ldots n-1}{1 \cdot 2 \cdot 3 \ldots n-1} \cdot 2^n \cdot x^n = \frac{2}{n} \frac{(2n-2)!}{(n-1)!(n-1)!} x^n == \frac{-2}{n} \binom{2n-2}{n-1}$$

The coefficient of x^n in $A(x)$ is the coefficient of x^n in $(1-(1-4x)^{1/2})/2$ which is thus

$$\frac{-1}{2} \cdot \frac{-2}{n} \binom{2n-2}{n-1} = \frac{1}{n} \binom{2n-2}{n-1}.$$

■

Exercise 414.

Catalan continued. Find the number of ways $2n$ points on a circle can be joined pairwise with n non-intersecting chords.

Solution.
Let a_n be the number of ways $2n$ points on a circle can be joined pairwise with n non-intersecting chords. Take a given point on the circle, let it be x. When x is joined by a chord to another point on the circle it divides the points on the circle into two sets of say $2i$ points on the left of the chord and $2n-2i-2$ points on the right (if we have an odd number of points on either side then we must have two intersecting chords). Summing up for all possible values of i we get the following recurrence for this problem:

$$a_n = \sum_{i=0}^{n-1} a_i a_{n-1-i}, \quad a_0 = a_1 = 1$$

Now we sum for all possible values of $n = 1,2,\ldots$ (note that we choose $a_0 = 1$, one can avoid this by writing separately the first and the last term of the sum which gives a_n above). Then we get:

$$\sum_{n=1}^{\infty} a_n x^n = \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} a_i a_{n-1-i} x^n \Rightarrow \; A(x) - a_0 = xA^2(x)$$

where $A(x)$ is the o.g.f. of $a_0, a_1, \ldots$. From the last equation we get that $A(x) = \frac{+1 \pm \sqrt{1-4x}}{2x}$. We choose only the negative root and from the Binomial Thm we get:

$$A(x) = \frac{1}{2x}(1 - \sum_{n\geq 0} \binom{\frac{1}{2}}{n} (-4x)^n) = \sum_{m\geq 0} \binom{\frac{1}{2}}{m+1} (-1)^m 2^{2m+1} z^m = \sum_m \frac{1}{m+1} \binom{2m}{m} z^m$$

Therefore, from the last expression we get that the coefficient of x^n in $A(x)$ is $\frac{1}{n+1}\binom{2n}{n}$. This is just the $n+1$-st Catalan number. One can avoid the recomputation of $A(x)$ by using $b_n = a_{n-1}$ and then the recurrence with respect to b_n follows.

■

Exercise 415.
Solve the following recurrence relation:

$$a_n = a_{n-1} + n^2 \ , \quad a_0 = 1$$

Solution.
Consider the answer to be of the form

$$a_n = An^3 + Bn^2 + Cn + D$$

We verify this and determine the coefficients, as follows.

$$\begin{aligned} a_n - a_{n-1} &= n^2 \\ A(n^3 - (n-1)^3) + B(n^2 - (n-1)^2) + C(n - (n-1)) &= n^2 \\ A(3n^2 - 3n + 1) + B(2n-1) + C &= n^2 \end{aligned}$$

This gives us the simultaneous equations

$$\begin{aligned} 3A &= 1 \\ -3A + 2B &= 0 \\ A - B + C &= 0 \end{aligned}$$

The solution for these is $A = 1/3$, $B = 1/2$, and $C = 1/6$. This gives the particular solution for the recurrence. To satisfy the initial condition, we must add the homogeneous solution $a_n = 1$. The final answer is

$$a_n = \frac{2n^3 + 3n^2 + n + 6}{6}$$

■

Exercise 416.

Solve the following recurrence relation:

$$a_n = a_{n-1} + \frac{1}{n(n+1)} \ , \quad a_0 = 1$$

Solution.

We can expand or not expand the sequence. We use the latter and an o.g.f. technique. Let $A(x) = \sum_{n=1}^{\infty} a_n x^n$ be the o.g.f. for a_n. Multiplying the recurrence by x^n for each $n \geq 1$ we get

$$\begin{aligned} a_1 x &= a_0 x + \frac{x}{1(1+1)} \\ a_2 x^2 &= a_1 x^2 + \frac{x^2}{2(2+1)} \\ &\vdots \end{aligned}$$

Summing, we get

$$\begin{aligned} A(x) - a_0 &= xA(x) + \sum_{n=1}^{\infty} \frac{x^n}{n(n+1)} \\ A(x) &= \frac{a_0 + \sum_{n=1}^{\infty} \frac{x^n}{n(n+1)}}{1-x} \\ A(x) &= (1 + \sum_{n=1}^{\infty} \frac{x^n}{n(n+1)})(1 + x + x^2 + \cdots) \end{aligned}$$

From here we can calculate the coefficient of x^n in the right hand side, which is

$$\begin{aligned} 1 + \sum_{i=1}^{n} \frac{1}{n(n+1)} &= 1 + \sum_{i=1}^{n} (\frac{1}{n} - \frac{1}{n+1}) \\ &= 1 + 1 - \frac{1}{n+1} \\ &= \frac{2n+1}{n+1} \end{aligned}$$

■

Exercise 417.
Solve the following divide and conquer recurrence relation exactly for the case where n is a power of two

$$S(n) = S(n/2) + 3n, \qquad S(1) = 1$$

Solution.
Let $n = 2^k$. Then from $S(n) = S(n/2) + 3n$, $S(1) = 1$ we get:

$$\begin{aligned} S(n) = S(2^k) &= S(2^{k-1}) + 32^k \\ &= S(2^{k-2}) + 32^{k-1} + 32^k \\ &= S(2^{k-3}) + 32^{k-2} + 32^{k-1} + 32^k \\ &= \vdots \\ &= S(2^0) + 3(2^1 + 2^2 + \ldots 2^k) \\ &= 1 + 3(2^1 + \ldots + 2^k) \\ &= 1 + 3 \cdot 2 \cdot (2^k - 1) = 6 \cdot 2^k - 5 = 6n - 5 \end{aligned}$$

■

Exercise 418.

Fibonacci revisited. For $a_0 = r, a_1 = s$, express in terms of the Fibonacci numbers ($f_{n+1} = f_n + f_{n-1}$, $f_0 = f_1 = 1$) the following recurrence relations.

$$a_{n+2} = a_{n+1} + a_n.$$

Then solve

$$a_{n+2} = a_{n+1} + a_n + c.$$

Solution.
(i) Let A(x) be the o.g.f. for a_n. Multiplying both sides of the recurrence by x^n, and summing over all values of n from 2 to ∞, we get

$$\begin{aligned}
\sum_{n=2}^{\infty} a_n x^n &= \sum_{n=2}^{\infty} a_{n-1}x^n + a_{n-2}x^n \\
A(x) - a_0 - a_1 x &= x(A(x) - a_0) + x^2 A(x) \\
A(x)(1 - x - x^2) &= x(s-r) + r \\
A(x) &= (x(s-r)+r)\frac{1}{1-x-x^2}
\end{aligned}$$

But that fraction is the o.g.f. for f_n. Thus

$$\begin{aligned}
a_n &= (s-r)f_{n-1} + rf_n \\
&= sf_{n-1} + rf_{n-2}
\end{aligned}$$

(ii) To use our result from (i), we can define $b_n = a_n + c$, and get a new recurrence relation:

$$b_n = b_{n-1} + b_{n-2}\ , \quad b_0 = r + c, \ \ b_1 = s + c$$

The solution to this is

$$\begin{aligned}
b_n &= (s+c)f_{n-1} + (r+c)f_{n-2} \\
&= cf_n + sf_{n-1} + rf_{n-2}
\end{aligned}$$

Converting back to a_n, we get our answer

$$a_n = (c-1)f_n + sf_{n-1} + rf_{n-2}$$

■

Exercise 419.
Solve

$$3a_n - 4a_{n-1} + a_{n-2} = 2^{-n} \ , \quad a_0 = 6, \ \ a_1 = 2.5$$

Solution.
We now solve the recurrence $3a_n - 4a_{n-1} + a_{n-2} = 2^{-n}$, $a_0 = 6$, $a_1 = 2.5$. The characteristic polynomial of this recurrence is : $3x^2 - 4x + 1 = 0$, with roots $x_{1,2} = 1, \frac{1}{3}$. Therefore the general solution of the homogeneous $3a_n - 4a_{n-1} + a_{n-2} = 0$ is

$$a_n^{gen} = c_1 \cdot 1^n + c_2 \cdot (\frac{1}{3})^n \ ,$$

c_1, c_2 constants to be computed later from the initial conditions. We now find a particular solution of the original solution. We try the following solution $a_n^{part} = A(\frac{1}{2})^n$ and we thus get from the original recurrence $3a_n - 4a_{n-1} + a_{n-2} = 2^{-n}$:

$$3A(1/2)^n - 4A(1/2)^{n-1} + A(1/2)^{n-2} = (1/2)^n$$

and by multiplying both sides with 2^n we finally get $A = -1$. Therefore a particulal solution to the recurrence is $a_n^{part} = (-1) \cdot (\frac{1}{2})^n$. A general solution to the original (non-homogeneous) one is the sum of the the two solutions i.e.

$$a_n = a_n^{gen} + a_n^{part} = c_1 \cdot 1^n + c_2 \cdot (\frac{1}{3})^n + (-1) \cdot (\frac{1}{2})^n$$

We now compute the constants c_1, c_2.

$$a_0 = 6 = c_1 \cdot 1 + c_2 \cdot 1 - 1 \Rightarrow \ c_1 + c_2 = 7$$

$$a_1 = 2.5 = c_1 + c_2 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{2} \Rightarrow \ 3c_1 + c_2 = 9$$

This system of equations gives $\begin{cases} c_1 = 1 \\ c_2 = 6 \end{cases}$ as a solution.
Therefore

$$a_n = 1 \cdot 1^n + 6 \cdot (\frac{1}{3})^n + (-1) \cdot (\frac{1}{2})^n$$

■

Exercise 420.

You are given the following system of recurrences. Solve for a_n and b_n.

$$\begin{aligned} a_n &= 2a_{n-1}+2b_n\ , \quad a_0=2 \\ b_n &= a_n+3b_{n-1}\ , \quad b_0=3 \end{aligned}$$

Solution.
If we use the o.g.f. of $a_n,\ b_n$ (let them be $A(x),\ B(x)$) we get that the system of equations

$$a_n = 2a_{n-1}+2b_n$$
$$b_n = a_n+3b_{n-1}$$

gives the following system of generating functions.

$$A(x)-a_0 = 2xA(x)+2(B(x)-b_0)$$
$$B(x)-b_0 = 3xB(x)+A(x)-a_0$$

Substituting $a_0=2,\ b_0=3$ we get:

$$A(x)(1-2x) = 2B(x)-4$$
$$B(x)(1-3x) = A(x)+1$$

Now, we solve the second equation with respect to $A(x)$ and replace the expression of $A(x)$ in terms of $B(x)$ in the first equation. We get:

$$B(x)(1-3x) = \frac{2B(x)}{(1-2x)}+\frac{-4}{(1-2x)}+1 \ \Rightarrow B(x)((1-3x)-\frac{2}{(1-2x)}) = \frac{-4}{(1-2x)}+1 \ \Rightarrow$$

$$B(x)\cdot\frac{6x^2-5x-1}{(1-2x)} = \frac{-4}{(1-2x)}+1 \ \Rightarrow B(x)\cdot\frac{(6x+1)(x-1)}{(1-2x)} = \frac{-4}{(1-2x)}+1 \ \Rightarrow$$

$$B(x) = \frac{-4}{(1+6x)(x-1)}+\frac{(-2x+1)}{(1+6x)(x-1)} \ \Rightarrow B(x) = \frac{4}{(1+6x)(1-x)}+\frac{(2x-1)}{(1+6x)(1-x)}$$

Now, we expand each of the two terms into simpler fractions. We thus get:

$$B(x) = \frac{\frac{24}{7}}{1+6x}+\frac{\frac{4}{7}}{1-x}+\frac{\frac{-8}{7}}{1+6x}+\frac{\frac{1}{7}}{1-x} \ \Rightarrow B(x) = \frac{\frac{16}{7}}{1+6x}+\frac{\frac{5}{7}}{1-x}$$

We know that $1/(1-ax)$ is the o.g.f. of a^n (and $1+6x = 1-(-6)x$). Therefore:

$$b_n = \frac{16}{7}(-6)^n+\frac{5}{7}1^n$$

Now, from $b_n = a_n+3b_{n-1}$ we have that $a_n = b_n-3b_{n-1}$ and therefore we get that:

$$a_n = \frac{24}{7}(-6)^n-\frac{10}{7}1^n$$

■

Exercise 421.

Solve the following recurrence relation:

$$4x_{n+1} - 5x_n + x_{n-1} = (0.5)^n \quad , \; x_0 = 2, \; x_1 = 1$$

Solution.
We derive the characteristic equation $4x^2 - 5x + 1 = 0$ from the recurrence:

$$4x_{n+1} - 5x_n + x_{n-1} = (0.5)^n \quad , \; x_0 = 2, \; x_1 = 1 \tag{6.4}$$

This equation has solutions $x_{1,2} = 1, \frac{1}{4}$. Therefore the general solution of the homogeneous equation given by (1) is:

$$x_n^{(hom)} = A \; 1^n + B \; (\frac{1}{4})^n$$

For a partial solution of the non-homogeneous of (1) we try $x_n = c(0.5)^n$. If we replace this in (1) and solve with respect to c we get that $c = -1$. Therefore a partial solution for the non-homogeneous of (1) (the initial conditions are ignored) is:

$$x_n^{(par)} = -(0.5)^n$$

and therefore, a general solution for (1) is given by:

$$x_n^{(gen)} = x_n^{(hom)} + x_n^{(par)} = A1^n + B(\frac{1}{4})^n - (0.5)^n \tag{6.5}$$

Substituting $n = 0$ and $n = 1$ in (2) and equating each case to the initial conditions $x_0 = 2$ and $x_1 = 1$, we get then that: $A + B - 1 = 2$ and $A + \frac{B}{4} - \frac{1}{2} = 1$ which give $A = 1$, $B = 2$. Therefore:

$$x_n^{(gen)} = 1 + 2(\frac{1}{4})^n - (0.5)^n \tag{6.6}$$

■

Exercise 422.

Solve the following system of recurrences:

$$x_{n+1} = 2x_n + 2y_n \quad , \; x_0 = 3$$

$$y_{n+1} = x_n + 3y_n \quad , \; y_0 = 6$$

Then solve the following recurrence:

$$w_{n+1} = \frac{2w_n + 2}{w_n + 3} \quad , \quad w_0 = 0.5$$

Solution.

Let $X(z)$ and $Y(z)$ be the o.g.f of the sequences x_n and y_n respectively. Then, from the system of recurrences:

$$x_{n+1} = 2x_n + 2y_n \quad , \; x_0 = 3$$

$$y_{n+1} = x_n + 3y_n \quad , \; y_0 = 6$$

we get, by summing up for all values of $n = 0, 1, 2, \ldots$ (and multiplying each equation by z^n) the following system of generating functions:

$$X(z) - x_0 = 2zX(z) + 2zY(z)$$

$$Y(z) - y_0 = zX(z) + 3zY(z)$$

From the first we get that $X(z) = \frac{3+2z}{(1-2z)}Y(z)$. Replacing this into the second one we get after some manipulations:

$$Y(z) = \frac{6(1-2z)}{(1-2z)(1-4z)} + \frac{3z}{(1-z)(1-4z)}$$

We note that:

$$\frac{6(1-2z)}{(1-2z)(1-4z)} = \frac{2}{(1-z)} + \frac{4}{(1-4z)}$$

and:

$$\frac{3z}{(1-z)(1-4z)} = \frac{-1}{(1-z)} + \frac{1}{(1-4z)}$$

and therefore:

$$Y(z) = \frac{1}{1-z} + \frac{5}{1-4z} \quad \Rightarrow \; y_n = 1^n + 5 \cdot 4^n$$

We replace $Y(z)$ in the expression we got for $X(z)$ and we similarly get that:

$$X(z) = \frac{-2}{1-z} + \frac{5}{1-4z} \; \Rightarrow \; x_n = -2 \cdot 1^n + 5 \cdot 4^n$$

For the second part identify that for $w_n = \frac{x_n}{y_n}$ the recurrence for w_n is equivalent to the system of the two recurrences above and a solution for w_n is easily derived. ■

Exercise 423.
Solve the following recurrence relation.

$$x_{n+1} - 6x_n + 9x_{n-1} = 2^n \ , \ x_0 = 3, \ x_1 = 10$$

Solution.
First guess a particular solution of the form $x'_n = A2^n$. Substituting into the recurrence this gives

$$\begin{aligned} A \cdot 2^{n+1} - 6A \cdot 2^n + 9A \cdot 2^{n-1} &= 2^n \\ 4A - 12A + 9A &= 2 \\ A &= 2 \\ x'_n &= 2^{n+1}. \end{aligned}$$

In order to find the homogeneous solution we observe that the characteristic polynomial is $\lambda^2 - 6\lambda + 9 = (\lambda - 3)^2$. So the solution is of the form $x_n = (Bn + C)3^n$ (recall this was how we handle a root of multiplicity more than one).
The complete solution is a combination of the two $x_n = x'_n + x_n = 2^{n+1} + (Bn + C)3^n$ where B and C are selected to satisfy the initial conditions. These give the pair of equations $3 = 2 + C$ and $10 = 4 + (B + C)3$, whose solutions are $B = C = 1$. Thus the final answer is as follows.

$$x_n = n3^n + 3^n + 2^{n+1}.$$

■

Exercise 424.
Solve the following recurrence relation:

$$x_{n+1} - x_n + x_{n-1} = 2^n \ , \ x_0 = 2, \ x_1 = 2$$

Solution.
The characteristic equation of the homogeneous is $a^2 - a + 1 = 0$ which gives two complex roots $a_{1,2} = \frac{1 \pm i\sqrt{3}}{2}$. Therefore, $x_n = Aa_1^n + Ba_2^n$. We try $x'_n = c2^n$ for a particular solution which gives $c = 2/3$. Then the general solution for our recurrence is:

$$x_n = x_n + x'_n = Aa_1^n + Ba_2^n + \frac{2}{3}2^n$$

We now apply the initial conditions to this solution. We get:

$$x_0 = 2 = A + B + \frac{2}{3}$$

and

$$x_1 = 2 = A(\frac{1 + i\sqrt{3}}{2}) + B(\frac{1 - i\sqrt{3}}{2}) + \frac{4}{3}$$

which give $A = B = \frac{2}{3}$. Therefore, the solution of the recurrence is:

$$x_n = \frac{2}{3}a_1^n + \frac{2}{3}a_2^n + \frac{4}{3}2^n = \frac{4}{3}\cos(\frac{n\pi}{3}) + \frac{2}{3}2^n$$

■

Exercise 425.

Hiding Fibonacci in plain sight. Solve the following recurrence relation:

$$a_n = a_{n-1} \cdot a_{n-2} \text{ , where } a_1 = 2, \text{ and } a_2 = 4.$$

Solution.
We take the logarithms base 2 of both sides. We thus have: $\lg a_n = \lg a_{n-1} + \lg a_{n-2}$. We then substitute $b_n = \lg a_n$. Then the given recurrence becomes $b_n = b_{n-1} + b_{n-2}$, where $b_1 = \lg a_1 = \lg 2 = 1$ and $b_2 = \lg a_2 = \lg 4 = 2$. This is just the Fibonacci recurrence. Let f_n be the n-th term of the Fibonacci recurrence. Then $a_n = 2^{f_n}$. ■

Exercise 426.

(a) Solve the following recurrence relation:

$$a_n - 7\ a_{n-1} + 16\ a_{n-2} - 12\ a_{n-3} = 0, \text{ where } a_0 = 2,\ a_1 = 7,\ a_2 = 21.$$

(b) Use some of the results of part (a) to solve the following recurrence relation:

$$a_n - 7\ a_{n-1} + 16\ a_{n-2} - 12\ a_{n-3} = 4^{n-1}, \text{ where } a_0 = 2,\ a_1 = 7,\ a_2 = 35.$$

Solution.

(a) The characteristic equation of the recurrence relation is $x^3 - 7x^2 + 16x - 12 = 0$, which has 3 roots, $x_{1,2} = 2$ (double root), and $x_3 = 3$. Therefore the general solution of the recurrence is:

$$a_n = A2^n + Bn2^n + C3^n \quad (1)$$

We substitute the initial conditions to (1) and solving a system of 3 equations on 3 unknows we easily find that $A = B = C = 1$, therefore the solution to the recurrence is $a_n = (n+1)2^n + 3^n$.

(b) We examine first the general solution a_n^{gen} of the homogeneous recurrence, that is: $a_n - 7a_{n-1} + 16a_{n-2} - 12a_{n-3} = 0$. This is identical to the one of part (a), i.e., $a_n^{gen} = A2^n + Bn2^n + C3^n$. We then find a partial solution of the non-homogeneous, that is:

$$a_n - 7a_{n-1} + 16a_{n-2} - 12a_{n-3} = 4^{n-1} \quad (2)$$

We fist try as a solution the $a_n^{par} = c4^n$, c a constant, to be computed by replacing a_n in (2) by a_n^{par}. Substituting that way in (2) and solving with respect to c we get that $c = 4$. Therefore, $a_n^{par} = 4 \cdot 4^n = 4^{n+1}$. We sum the two partial results to find the general solution of the original recurrence relation:

$$a_n = a_n^{gen} + a_n^{par} = A2^n + Bn2^n + C3^n + 4^{n+1}$$

Now and only now can we apply the initial conditions. We get a system of 3 equations with 3 unknowns that gives as a solution the $A = C = -1$, $B = -2$, and thus $a_n = (-1-2n)2^n - 3^n + 4^{n+1}$. ■

Exercise 427.

Derangements revisited. Let D_n denote the number of derangements of n numbers. Let $D(x)$ be the exponential generating function of D_n.
(a) Show that:

$$n! = \sum_k \binom{n}{k} D_{n-k} \quad ,(\ n \geq 0\ )$$

(b) Show that $D(x) = \frac{e^{-x}}{1-x}$, and hence deduce that:

$$D_n = n!\ (1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + (-1)^n \frac{1}{n!})$$

Solution.
(a) The number of permutation over n numbers is just $n!$. We count this number by counting all permutations that have exactly k fixed points, for every $k = 0, \ldots, n$, and then we sum up all such counts. Suppose k specific numbers will be fixed points in a permutation. Then the other $n-k$ numbers must form a derangement on $n-k$ numbers in this permutation, to have only k fixed points in it. We can choose the k fixed points in $\binom{n}{k}$ ways and for each such choice we have D_{n-k} derangements on the other $n-k$ numbers. Thus we get: $n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k}$.
b) We now multiply both sides of the previous equation by $x^n/n!$, and summing for all values of n we get:

$$\sum_{n=0}^{\infty} n! \frac{x^n}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} D_{n-k} \cdot 1^k \cdot \frac{x^n}{n!}$$

The left hand side is just $1/(1-x)$, while the right hand side (from the formula that gives the coefficient of $x^n/n!$ of the product of two e.g.f.'s) is just e^x (this is the e.g.f. of the sequence $a_i = 1, \forall i$) times $D(x)$, the e.g.f of D_n. Thus $1/(1-x) = e^x D(x)$, and $D(x) = e^{-x}/(1-x)$.
On a separate note, suppose that the sequence $f_n, n \geq 0$ has o.g.f $F(x)$. Then $F(x)/(1-x)$ is the o.g.f. of the sequence $\sum_{i=0}^{n} f_i$, $n \geq 0$. The proof is straightforward from the fact that $F(x)/(1-x) = (f_0 + f_1 x + f_2 x^2 + \ldots)(1 + x + x^2 + \ldots)$. We also note that the coefficient of $x^n/n!$ in $e^{-x}/(1-x)$ is $n!$ times the coefficient of x^n in $e^{-x}/(1-x)$.
Since $D(x) = e^{-x}/(1-x)$, then the coefficient of x^n in $D(x)$ is just (from the result of the previous paragraph) $\sum_{i=0}^{n} (-1)^i/i!$. Therefore the coefficient of $x^n/n!$ in $D(x)$ is $n!$ times the previous sum, that is $n!(1 - 1/1! + 1/2! - \ldots + (-1)^n/n!)$, as desired. ■

Exercise 428.

More on derangements. Let A be a set of n objects $\{1,2,\ldots,n\}$. Let P be the set of all permutations on those objects, and let P_i be the property that object i is in poisition i of a permutation. Then $D(n)$ is the number of permutations of P attesting NONE of P_i. Note that $A(i) = |P_i| = (n-1)!$. Likewise, $A(i,j) = (n-2)!$ and so on. Derive the expression for the number of derangements on n objects.

Solution.
Thus utilizing the inclusion-exclusion formula obtained earlier, i.e.

$$|A| - N = |A| - \sum_i A(i) + (-1)^2 \sum_{i<j} A(i,j) + (-1)^3 \sum_{i<j<k} A(i,j,k) + \ldots$$

$$\begin{aligned}
D_n &= |P| - \binom{n}{1} A(i) + \binom{n}{2} A(i,j) + \ldots + (-1)^n \binom{n}{n} A(1,2,\ldots,n) \\
&= n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! + \ldots + (-1)^j \binom{n}{j}(n-j)! + \ldots \\
&= n! - \frac{n!}{1!} + \frac{n!}{2!} + \ldots + (-1)^j \frac{n!}{j!} + \ldots \\
&= n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots\right) \approx n!/e.
\end{aligned}$$

■

Exercise 429.
Find the number of surjections for function $f : A \to B$, where $|A| = m$ and $|B| = n$. (Function f is surjective if $Im(f) = B$.

Solution.
Let the elements of B be $B = \{b_1, \ldots b_n\}$. Let $P(i)$ be the property that b_i is not in the image of f. Let A be the set of all functions from A to B i.e. we have m^n such functions. Then $A(i) = (m-1)^n$. Likewise $A(i,j,k) = (m-3)^n$ and so on. The number of surjections s_n from A to B is then the number of functions of A having no property $P(i)$ i.e.

$$\begin{aligned} s_n &= |A| - \sum_i A(i) + (-1)^2 \sum_{i<j} A(i,j) + (-1)^3 \sum_{i<j<k} A(i,j,k) + \ldots \\ &= m^n - \binom{m}{1}(m-1)^n + \binom{m}{2}(m-2)^n + \ldots + (-1)^m \binom{m}{m}(m-m)^n \\ &= \sum_{i=0}^{m} (-1)^i \binom{m}{i}(m-i)^n. \end{aligned}$$

Note that if $n < m$ the sum is 0. ■

Exercise 430.

A latin square of order n is an $n \times n$ matrix whose every row and column contains the n natural numbers $\{1, 2, \ldots, n\}$ exactly once.
Think of have $2n$ teams in a sport and one arranges over $2n-1$ game DAYS for one team to play against any other team.

$$A = \begin{bmatrix} 4 & 1 & 2 & 3 \\ & 4 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}$$

When do teams play a game against team 1? Fill the remaining schedule.

Solution.
Ignoring the diagonal of A, $a_{12} = 1$ to indicate that team 1 and team 2 play on the 1st game DAY. Likewise team 1 and team 3 play on the 2nd game DAY, and team 1 and team 4 play on the 3nd game DAY. In other words $a_{ij} = k$ if team i and team j play a game on the k-th day.

$$A = \begin{bmatrix} 4 & 1 & 2 & 3 \\ 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & 2 \\ 3 & 2 & 1 & 4 \end{bmatrix}$$

This schedule, though a latin square has a problem: team 1 and 2 play the first week but also the second week. Thus we need to have a Latin square that is symmetric. The following makes more sense.

$$A = \begin{bmatrix} 4 & 1 & 2 & 3 \\ 1 & 4 & 3 & 2 \\ 2 & 3 & 4 & 1 \\ 3 & 2 & 1 & 4 \end{bmatrix}$$

■

Exercise 431.

(a) Let A and B be two $n \times n$ matrices and the join(A,B) is a matrix C also $n \times n$ whose $c_{ij} = (a_{ij}, b_{ij})$. Two latin squares A, B are orthogonal if all the entries of C are distinct. $R(n)$ is the maximum number of Latin Squares that are mutually orthogonal. Show that $R(n) \leq n-1$ for $n \geq 2$.
(b) Show then that $R(p) = p-1$ for all prime p.

Solution.
Let A_k, $k = 1, \ldots, m$ be m mutually orthogonal latin squares of dimension n. Relabel them so that first row of each one is $1, 2, \ldots, n$. Consider the element of row 2 and column 1, i.e. $(2,1)$. Since every square relabeled has a 1 in $(1,1)$, there are only $n-1$ choices for $(2,1)$. But no two squares can have say 5 in $(2,1)$ since otherwise their join would have $(5,5)$ in $(2,1)$ an we also know that the join at $(1,5)$ is also $(5,5)$ thus violating the orthogonality condition. So $m \leq n-1$.
Let A_k be p Latin square $k = 1, \ldots, p$ that are mutually orthogonal latin squares of dimension n. Let for A_k, $a_{ij}^{(k)} = k * i + j \pmod{p}$. We show that A_k is a latin square. Pick two elements of the same row i, call them at columns l, m but for different columns $l \neq m$.

$$k * i + l \equiv k * i + m \pmod{p}$$

implies $l - m \equiv 0 \pmod{p}$. As l, m are indexed columns, this is equivalent to $l = m$. Repeat the same for column i and rows l, m then $l \neq m$.

$$k * l + i \equiv k * m + i \pmod{p}$$

This implies

$$k(l-m) \equiv 0 \pmod{p}$$

Then p, a prime number, divides $k(l-m)$, but $k < p$ and thus p divides $l - m$ that is $l \equiv m \pmod{p}$. Given that both l, m are smaller than p this implies $l = m$ contradicting $l \neq m$. We now show that two latin square A_r and A_s are mutually orthogonal. A similar argument can be used then.
Magic Squares. One arranges of $n \times n$ matrix A each of the integers $1, 2, \ldots, n^2$ so that the sum of a row or column or the two main diagonals is equal to $n(n^2+1)/2$.

$$\begin{bmatrix} 6 & 1 & 8 \\ 7 & 5 & 3 \\ 2 & 9 & 4 \end{bmatrix}$$

■

Exercise 432.

Let A_i, $i = 1,\ldots,N$ be a collection of sets. Let r_i, $i = 1,\ldots,N$ be a collection of a system of distinct representatives of A_i, i.e. $r_i \in A_i$ and r_i distinct.
Condition: Show that for all $i \leq N$, the union of j of the sets A_i contains at least j elements.

Solution.
If there is an SDR for A_i, then clearly then the condition above is clearly satisfied. No consider that the condition is satisfied. We shall show that there is an SDR.
We starting removing elements from A_i so that the condition is still satisfied and thus we get $B_i \subseteq A_i$ for which the removal of an element would break the condition above. It will then be shown that $c(B_i) = 1$ and all sets are disjoing thus the DR of B_i would be its only element.
Suppose that a B_l has two elements c,d. If any one of the two gets removed from B_l then the condition breaks down. Thus there are sets C,D such that

$$C_c = (B_l - c) \cup \cup_{i \in C} B_i, D_d = (B_l - d) \cup \cup_{i \in D} B_i.$$

Then if the condition is violated $c(C_c) < c(C) + 1$ and $c(D_d) < c(D) + 1$. But then

$$C_c \cup D_d = B_l \cup \cup_{i \in C \cap D} B_i$$

and

$$\cup_{i \in C \cap D} B_i \subseteq C_c \cap D_d$$

The by the condition we have,

$$c(C_c \cup D_d)$$

■

Part III

Analysis of algorithms tools

Chapter 7

Asymptotic notation and other recurrence relations

7.1 Short review on asymptotic notation

7.1.1 Big-oh O notation

Definition 7.1 **Big-Oh O notation**

Let $f(n)$ and $g(n)$ be asymptotically non-negative functions.
For a function $g(n)$ we denote by $O(g(n))$ the set of all functions $f(n)$ that have the following property.
There exist positive constants c_2, n_0 such that

$$\exists \textbf{pos.const } c_2, n_0 : \quad 0 \leq f(n) \leq c_2 g(n) \quad \forall n \geq n_0$$

For a function $f(n)$ that satisfies this property we write

$$f(n) = O(g(n)) \quad \textbf{or equivalently} \quad f(n) \in O(g(n)).$$

In other words,

Definition 7.2 **Limit based O**

If the limit $\lim_{n\to\infty} f(n)/g(n)$ exists,

$$f(n) = O(g(n)) \Leftrightarrow \exists d_2, \text{PosConst}(d_2) : \lim_{n\to\infty} \frac{f(n)}{g(n)} \leq d_2$$

7.1.2 Big-Omega Ω notation

Definition 7.3 **Big-Omega Ω notation**

Let $f(n)$ and $g(n)$ be asymptotically non-negative functions.
For a function $g(n)$ we denote by $\Omega(g(n))$ the set of all functions $f(n)$ that have the following property.
There exist positive constants c_1, n_0 such that

$$\exists \textbf{pos.const } c_1, n_0 : \quad 0 \leq c_1 g(n) \leq f(n) \quad \forall n \geq n_0$$

For a function $f(n)$ that satisfies this property we write

$$f(n) = \Omega(g(n)) \quad \textbf{or equivalently} \quad f(n) \in \Omega(g(n)).$$

In other words,

Definition 7.4 **Limit based Ω**

If the limit $\lim_{n\to\infty} g(n)/f(n)$ exists,

$$f(n) = \Omega(g(n)) \Leftrightarrow \exists d_1, \text{PosConst}(d_1) : \lim_{n\to\infty} \frac{g(n)}{f(n)} \leq d_1$$

7.1.3 Theta Θ notation

Definition 7.5 **Theta Θ notation**

Let $f(n)$ and $g(n)$ be asymptotically non-negative functions.
For a function $g(n)$ we denote by $\Theta(g(n))$ the set of all functions $f(n)$ that have the following property. There exist positive constants c_1, c_2, n_0 such that

$$\exists \textbf{pos.const } c_1, c_2, n_0 : \quad 0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n) \quad \forall n \geq n_0$$

For a function $f(n)$ that satisfies this property we write

$$f(n) = \Theta(g(n)) \quad \textbf{or equivalently} \quad f(n) \in \Theta(g(n)).$$

In other words,

Definition 7.6 **O and Ω based Θ notation**

$f(n) = \Theta(g(n))$ if and only if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

7.1.4 Little-oh o notation

Definition 7.7 **Little-oh o notation**

Let $f(n)$ and $g(n)$ be asymptotically non-negative functions.

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0 \Rightarrow f(n) = o(g(n)).$$

7.1.5 Little-omega ω notation

Definition 7.8 **Little-omega ω notation**

Let $f(n)$ and $g(n)$ be asymptotically non-negative functions.

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty \Rightarrow f(n) = \omega(g(n)).$$

Definition 7.9 **Quick o, ω, Θ**

In order to asymptotically compare two non-negative functions $f(n)$ and $g(n)$ with the three symbols o, ω, Θ, we calculate first $\lim_{n\to\infty} \frac{f(n)}{g(n)}$ and then determine the case applicable as follows. (Parameter c is a positive constant.)

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \begin{cases} 0 & \Rightarrow f(n) = o(g(n)) \\ \infty & \Rightarrow f(n) = \omega(g(n)) \\ c \neq 0 & \Rightarrow f(n) = \Theta(g(n)) \end{cases}$$

Corollary 7.1 **$o \Rightarrow O$ implication**

If $f(n) = o(g(n))$, then $f(n) = O(g(n))$.

Corollary 7.2 **$\omega \Rightarrow \Omega$ implication**

If $f(n) = \omega(g(n))$, then $f(n) = \Omega(g(n))$.

Corollary 7.3 **$\Theta \Leftrightarrow \Omega \bigwedge O$**

$f(n) = \Theta(g(n))$, if and only if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$.

7.2 Warm-up : limits

Exercise 433.
Find the following limits.

(i)

$$\lim_{n\to\infty} n^2/2^n,$$

(ii)

$$\lim_{n\to\infty} 2^{\lg n}/n,$$

(iii)

$$\lim_{n\to\infty} n!/2^n.$$

Note $n! = 1 \cdot 2 \cdot \ldots \cdot n$.

Solution.
$\lim_{n\to\infty} \frac{n^2}{2^n} = 0$, after two applications of L'Hopital's theorem.
$\lim_{n\to\infty} 2^{\lg n}/n = \lim_{n\to\infty} n/n = 1$.
Note that from Stirling's approximation formula, $n! \approx (n/e)^n$.
$\lim_{n\to\infty} n!/2^n = \lim_{n\to\infty}(n/2)\cdots(n-1)/2\ldots(3/2)\cdot(2/2)\cdot(1/2) = \infty$, or
$\lim_{n\to\infty} n!/2^n = \lim_{n\to\infty}(n/e)^n/2^n = \lim_{n\to\infty}(n/(2e))^n = \infty$ ■

Exercise 434.
Find the following limit.

$$\lim_{n\to\infty} \lg(n!)/n\lg n,$$

Solution.
Note that from Stirling's approximation formula, $n! \approx (n/e)^n$, we have

$$\lim_{n\to\infty} \frac{\lg(n!)}{n\lg n} = \lim_{n\to\infty} \frac{\lg(n/e)^n}{n\lg n} = \lim_{n\to\infty} \frac{n\lg n - n\lg e}{n\lg n} = 1.$$

■

7.3 Asymptotic notation

Exercise 435.
TRUE or FALSE. Any other response is wrong.

(1) $2^{\lg n} = \Theta(n^2)$.
(2) $n/\lg n = \Theta(\lg(n!))$.
(3) $3^n = o(4^{n/2})$.
(4) $n! = \Theta((n+1)!)$.
(5) $2^{\lg n} = o(n^2)$.
(6) $n \lg n = \Theta(\lg(n!))$.
(7) $2^n = o(3^n)$.
(8) $n! = \Theta((n-1)!)$.

Solution.

(1) $2^{\lg n} = \Theta(n^2)$. **FALSE**
(2) $n/\lg n = \Theta(\lg(n!))$. **FALSE**
(3) $3^n = o(4^{n/2})$. **FALSE**
(4) $n! = \Theta((n+1)!)$. **FALSE**
(5) $2^{\lg n} = o(n^2)$. **TRUE**
(6) $n \lg n = \Theta(\lg(n!))$. **TRUE**
(7) $2^n = o(3^n)$. **TRUE**
(8) $n! = \Theta((n-1)!)$. **FALSE**

■

Exercise 436.
Answer the following questions as indicated with a ?. In other word replace ? with o, Θ or ω.

1. $10 = ?(1)$.
2. $n = ?(\lg n)$.
3. $2^{\lg n} = ?(n)$.
4. $2^n = ?(n)$.
5. $n = ?(n^3)$.
6. $2^n = ?(3^n)$.
7. $n \lg n = ?(\lg(n!))$.
8. $2^n = ?(n^2)$.
9. $n! = ?(2^n)$.
10. $n^{2/\lg n} = ?(n)$.
11. $n! = ?(n)$.
12. $n^3 = ?(2^{\lg n})$.

Solution.

1. Since $\lim 10/1 = 10$ as $n \to \infty$, we have $10 = \Theta(1)$.
2. Since $\lim n/\lg n = \lim 1/(1/n) = \infty$ as $n \to \infty$, we have $n = \omega(\lg n)$.
3. Since $\lim 2^{\lg n}/n = \lim n/n = 1$ as $n \to \infty$, we have $2^{\lg n} = \Theta(n)$.
4. Since $\lim 2^n/n = \infty$ as $n \to \infty$, we have $2^n = \omega(n)$.
5. Since $\lim n/n^3 = 0$ as $n \to \infty$, we have $n = o(n^3)$.
6. Since $\lim 2^n/3^n = 0$ as $n \to \infty$, we have $2^n = o(3^n)$.
7. Note that by Stirling's approximation formula $\lg(n!) \approx n \lg n$. Since $\lim n \lg n/\lg(n!) = c > 0$ as $n \to \infty$, and c some constant, we have $n \lg n = \Theta(\lg(n!))$.
8. $\lim 2^n/n^2 = \infty$ as $n \to \infty$, we have $2^n = \omega(n^2)$.
9. $\lim n!/2^n = \infty$ as $n \to \infty$, we have $n! = \omega(2^n)$.
10. $$\frac{n^{2/\lg n}}{n} = \frac{(2^{\lg n})^{2/\lg n}}{n} = \frac{2^2}{n} = \frac{4}{n} \to 0$$
    Thus $n^{2/\lg n} = o(n)$.
11. From Stirling approximation formula, $n! \approx (n/e)^n$. Therefore, $\frac{n!}{n} \to \infty$. Thus $n! = \omega(n)$.
12. We have that $2^{\lg n} = n$. By an earlier example, we have $n^3 = \omega(n)$, and thus $n^3 = \Omega(n)$.

■

Exercise 437.

Replace ? with the tightest of o, ω, Θ that best describes the relationship between the two sides. If necessary, use the approximation $n! \approx (n/e)^n$, where $e = 2.71\ldots$.

(1) $\lg(n!) = ?\ (n \lg n)$.
(2) $\lg(n!) = ?\ (n^2)$.
(3) $n^n = ?\ (n!)$.
(4) $2^n = ?\ (n!)$.
(5) $2^{\lg n} = ?\ (n)$.
(6) $n/\lg n = ?\ (n)$.
(7) $2^{3\lg n} = ?\ (n^3)$.
(8) $2^n = ?\ (3^n)$.
(9) $(n!) = ?\ ((n+1)!)$.
(10) $2^{\lg n} = ?\ (n^3)$.
(11) $n \lg n = ?\ (\lg(n!))$.
(12) $4^{2n} = ?\ (16^{n/2})$.
(13) $2^{\lg n} = ?\ (n^2)$.
(14) $2^{610n} = ?\ (610^{2n})$.
(15) $2^{n+n} = ?\ (2^{610n})$.
(16) $(n!) = ?\ ((n+610)^n)$.
(17) $3^{2n} = ?\ (2^{3n})$.

Solution.
From Stirling's approximation formula for the factorial $n! \approx (n/e)^n$. This answers the first two questions. Properties of the factorial answer the next two. The definition of $\lg n$ leads to the last answer.

(1) $\lg(n!) = \Theta(n \lg n)$.
(2) $\lg(n!) = o(n^2)$.
(3) $n^n = \omega(n!)$.
(4) $2^n = o(n!)$.
(5) $2^{\lg n} = \Theta(n)$.
(6) $n/\lg n = o(n)$.
(7) $2^{3\lg n} = \Theta(n^3)$.
(8) $2^n = o(3^n)$.
(9) $(n!) = o((n+1)!)$.
(10) $2^{\lg n} = o(n^3)$.
(11) $n \lg n = \Theta(\lg(n!))$.
(12) $4^{2n} = \omega(16^{n/2})$.
(13) $2^{\lg n} = o(n^2)$.
(14) $2^{610n} = \omega(610^{2n})$.
(15) $2^{n+n} = o(2^{610n})$.
(16) $(n!) = o((n+610)^n)$.
(17) $3^{2n} = \omega(2^{3n})$.

■

Exercise 438.
Rank the following functions by order of growth; that is, find an arrangement $g_1, g_2, g_3, \ldots$ of the functions satisfying $g_1 = \Omega(g_2), \quad g_2 = \Omega(g_3), \quad g_3 = \Omega(g_4), \quad \ldots$. Partition your list in equivalence classes such that $g(n)$ and $h(n)$ are in the same class if and only if $h(n) = \Theta(g(n))$.
For example for functions $\lg n, n, n^2$, and $2^{\lg n}$ you could write: $\quad n^2, \quad \{n, 2^{\lg n}\}, \quad \lg n$

$$n,\ (n-2)!,\ 5^n,\ (1/3)^n,\ (\lg n)^{\lg n},\ 4^{\lg n},\ n\lg n,\ \lg(n!),\ 2^{\lg n},\ n^2,\ 2^{4\lg n},\ 5,\ n^{\lg\lg n},\ n!,\ 3^n,\ n^n$$

Solution.
Use the results of prior problems to conclude the following.

$$n^n,\ n!,\ (n-2)!,\ 5^n,\ 3^n,\ \left\{(\lg n)^{\lg n}, n^{\lg n}),\right\}\ 2^{4\lg n},\ \left\{4^{\lg n}, n^2)\right\},\ \{n\lg n, \lg(n!)\},\ \left\{2^{\lg n}, n\right\}\ 5,\ (1/3)^n.$$

■

Exercise 439.
Rank the following functions by order of growth; that is, find an arrangement $g_1, g_2, g_3, \ldots$ of the functions satisfying $g_1 = \Omega(g_2), \quad g_2 = \Omega(g_3), \quad g_3 = \Omega(g_4), \quad \ldots$. Partition your list in equivalence classes such that $g(n)$ and $h(n)$ are in the same class if and only if $h(n) = \Theta(g(n))$.
For example for functions $\lg n, n, n^2$, and $2^{\lg n}$ you could write: $\quad n^2, \quad \{n, 2^{\lg n}\}, \quad \lg n$

$$e^n, (n+1)!, 2^{2^n}, 2^{2^{n+1}}, n!, n2^n, 2^n, (3/2)^n, n^{\lg\lg n}, (\lg n)^{\lg n}), \sqrt{2})^{\lg n}, (\lg n)!, n^2, n^3, 4^{\lg n}, n\lg n, n, \lg(n!), 2^{\lg n}.$$

Solution.

$$2^{2^{n+1}}, 2^{2^n}, (n+1)!, n!, e^n, n2^n, 2^n, (3/2)^n, (n^{\lg\lg n}, (\lg n)^{\lg n}), (\lg n)!, n^3, (n^2, 4^{\lg n}), (n\lg n, \lg(n!)), n = 2^{\lg n}, (\sqrt{2})^{\lg n},$$

■

Exercise 440.
Let $a_i > 0$ be constant for all i. Then

$$p(n) = a_0 + a_1\, n + \ldots + a_d\, n^d$$

Show that $p(n) \leq 2Cn^d$, for $n > 1$, where $C = \max\{a_0, a_1, \ldots a_d\}$. Show then that $p(n) = O(n^d)$.

Solution.
We have that $p(n) = a_0 + a_1\, n + \ldots + a_d\, n^d$. Let $C = \max\{a_0, a_1, \ldots a_d\}$. As at least $a_d > 0$ is positive and all a_i are constant, we conclude that $C > 0$ and constant.

For $n \geq 2$ and using the fact that $1 + n + \ldots + n^{d-1} = (n^d - 1)/(n-1)$, $(n \neq 1)$ we have that

$$\begin{aligned}
p(n) &= a_0 + a_1\, n + \ldots + a_{d-1}\, n^{d-1} + a_d\, n^d \\
&\leq C + Cn + \ldots + Cn^{d-1} + C\, n^d \\
&= C(1 + n + \ldots + n^{d-1}) + C\, n^d \\
&= C\frac{n^d - 1}{n-1} + C\, n^d \\
&\leq C\frac{n^d - 1}{1} + C\, n^d \\
&\leq C(n^d - 1) + C\, n^d \\
&\leq Cn^d + C\, n^d \\
&= 2Cn^d
\end{aligned}$$

Thus for $c_2 = 2C$ and $n_0 = 2$ we have $p(n) = O(n^d)$. ■

Exercise 441.

Show that $n^2 - 2n = \Theta(n^2)$. Provide constants c_1, c_2, n_0.

Solution.
We determine POSITIVE CONSTANTS c_1, c_2, n_0 so that for all $n \geq n_0$.

$$c_1 n^2 \leq n^2 - 2n \leq c_2 n^2$$

We separate the O from the Ω part and then combine the two parts.
Part O first. As $n^2 - 2n \leq n^2$ for all $n \geq 1$, we have shown that $n^2 - 2n \leq c_2\, n^2$, for $c_2 = 1$ and for all $n \geq n_1 = 1$. This is equivalent to showing that $n^2 - 2n = O(n^2)$ as well.
Part Ω next. As $n^2/2 \leq n^2 - 2n$ for all $n \geq 4$, we have shown that $c_1 n^2 \leq n^2 - 2n$, for $c_1 = 1/2$ and $n \geq n_2 = 4$. This is equivalent to showing that $n^2 - 2n = \Omega(n^2)$ as well.
Combine the two. The n_0 in the definition is the largest of n_1 and n_2 i.e. $n_0 = 4$. Then, for $c_1 = 1/2$, $c_2 = 1$ and for all $n \geq n_0 = 4$, we have

$$c_1 n^2 \leq n^2 - 2n \leq c_2 n^2$$

■

Exercise 442.
Is $2^n = 2^{\Theta(n)}$?

Solution.
No it is not. Example: $g(n) = 2n = \Theta(n)$. But $2^n = o(2^{g(n)})$, as $2^n/2^{2n} \to 0$. ■

Exercise 443.
Show that $n^2 - 10n + 2 = O(n^2)$. Use the definition of Ø.

Solution.
We are going to determine for $f(n) = n^2 - 10n + 2$ and $g(n) = n^2$, that there

$$\exists \text{ positive constants } c_2, n_0 : \ f(n) \leq c_2 g(n) \ \forall\, n \geq n_0.$$

Towards this

$$f(n) = n^2 - 10n + 2 \leq n^2 + 0 + 2 \leq n^2 + 0 + 2n^2 \leq 3n^2$$

Note that $-10n \leq 0$ is true for all $n \geq 0$ and thus for positive n_0 and $n \geq n_0$. Moreover $2 \leq 2n^2$ for all $n \geq 1$ and this requires $n_0 \geq 1$.
Conclusion. There exist $c_2 = 3$ and $n_0 = 1$ such that $f(n) = n^2 - 10n + 2 = O(n^2)$ for all $n \geq n_0 = 1$. ■

Exercise 444.
Show that $n^2 - 10n + 2 = \Omega(n^2)$. Use the definition of Ω.

Solution.
We are going to determine for $f(n) = n^2 - 10n + 2$ and $g(n) = n^2$, that there

$$\exists \text{ positive constants } c_1, n_0 : \ 0 \leq c_1 g(n) \leq f(n) \ \forall\, n \geq n_0.$$

Towards this

$$\frac{1}{2}n^2 \leq n^2 - \frac{1}{2}n^2 \leq n^2 - 10n \leq n^2 - 10n + 0 \leq n^2 - 10n + 2.$$

For $n^2 - 10n \geq n^2 - (1/2)n^2$ to hold we need $n^2 \geq 20n$ i.e. $n \geq 20$. Thus $n_0 \geq 20$; we pick $n_0 = 20$.
Conclusion. There exist $c_1 = 1/2$ and $n_0 = 20$ such that $f(n) = n^2 - 10n + 2 = \Omega(n^2)$ for all $n \geq n_0 = 20$.
We can also show that $n^2 - 10n + 2 = \Theta(n^2)$. This is true for $c_1 = 1/2$ and $c_2 = 3$ and $n_0 = \max\{20, 1\} = 20$. ■

Exercise 445.
Show that

$$\sum_{i=1}^{n} i^2 = \Theta(n^3).$$

What are the values of c_1, c_2 and n_0? Justify your answer.

Solution.
Omitted. ■

Exercise 446.

Use the definitions to show that $n^5 - 25n = \Theta(n^5)$.

Solution.
We show first that $n^5 - 25 = O(n^5)$ and then that $n^5 - 25 = \Omega(n^5)$.

Case a. Show that $n^5 - 25 = O(n^5)$.
For all $n \geq 1$ we have that

$$n^5 - 25 \quad \leq \quad n^5$$

Therefore there exist constant $n_2 = 1$ and $c_2 = 1$ such that $n^5 - 25 \leq c_2 n^5$ for all $n \geq n_2$. This proves the claim.
Technique 1. What we use in this proof is the fact that $n^2 \pm An \pm B$ is bounded above, for positive A, B, by $n^2 + An^2 + Bn^2 \leq (1 + A + B)n^2$.
Case b. Show that $n^5 - 25 = \Omega(n^5)$

Technique 1 can not be used in this case. The next step is non-trivial. We bound $n^5 - 25$ from below by $n^5/2$. This is so as long as

$$\begin{aligned} n^5 - 25 &\geq n^5/2 \Leftrightarrow \\ n^5/2 &\geq 25 \Leftrightarrow \\ n^5 &\geq 50 \Leftrightarrow \\ n &\geq 3 \end{aligned}$$

We can do this as long as n is not zero; this is true since for all cases we assume that at least $n \geq 1$. Therefore there exist constant $n_1 = 3$ and $c_1 = 1/2$ such that $n^5 - 25 \geq n^5/2$ for all $n \geq n_1 = 3$. This proves the claim. Note that 3 is not the best possible constant. We don't need to find the best possible constant but THE EASIEST POSSIBLE!
In order to show that $n^5 - 25 = \Theta(n^5)$ we need to establish c_1, c_2 and n_0. c_1 and c_2 are $1/2$ and 1 respectively. $n_0 = \max(n_1, n_2) = 3$. For these values the problem is thus shown. ■

Exercise 447.

(a) Show that $(n-3)^2 = \Theta(n^2)$ by providing the n_0, c_1, c_2 in the definition of the Θ notation.
(b) Show that $(n+3)^2 = \Theta(n^2)$ by providing the n_0, c_1, c_2 in the definition of the Θ notation.
(c) Show that $n^5 - 18n^3 = \Theta(n^5)$ by providing the n_0, c_1, c_2 in the definition of the Θ notation.

Solution.
Part (a): We show first that $(n-3)^2 = O(n^2)$ and then that $(n-3)^2 = \Omega(n^2)$.
Case 1. Show that $(n-3)^2 = O(n^2)$
For all $n \geq 1$ we have that

$$\begin{aligned}(n-3)^2 &= n^2 - 6n + 9 \\ &\leq n^2 + 0 \cdot n^2 + 9 \\ &\leq n^2 + 0 \cdot n^2 + 9 \cdot n^2 \\ &\leq 10n^2\end{aligned}$$

Therefore there exist constant $n_2 = 1$ and $c_2 = 10$ such that $(n-3)^2 \leq c_2 n^2$ for all $n \geq n_2$. This proves Claim 1.
Technique 1. We upper bound $n^2 \pm An \pm B$, for positive A, B as follows: $n^2 \pm An \pm B \leq (1 + A + B)n^2$.

Case 2. Show that $(n-3)^2 = \Omega(n^2)$
Technique 1 can not be used in this case. However, since $9 \geq 0$ we can eliminate 9 easily.

$$\begin{aligned}(n-3)^2 &= n^2 - 6n + 9 \\ &\geq n^2 - 6n\end{aligned}$$

The next step is non-trivial. We bound $n^2 - 6n$ from below by $n^2/2$. This is so as long as

$$\begin{aligned}n^2 - 6n &\geq n^2/2 \Leftrightarrow \\ n^2/2 &\geq 6n \Leftrightarrow \\ n/2 &\geq 6 \Leftrightarrow \\ n &\geq 12\end{aligned}$$

We can do this as long as n is not zero; this is true since for all cases we assume that at least $n \geq 1$. Therefore there exist constant $n_1 = 12$ and $c_1 = 1/2$ such that $(n-3)^2 \geq n^2/2$ for all $n \geq n_1 = 12$. This proves Claim 2. In order to show that $(n-3)^2 = \Theta(n^2)$ we need to establish c_1, c_2 and n_0 by combining the results of Claim 1 and Claim 2. Pick $c_1 = 1/2$ and $c_2 = 10$ and $n_0 = 12$.

(b) (O part first) To establish an upper bound we observe that,

$$(n+3)^2 = n^2 + 6n + 9 \leq n^2 + 6n^2 + 9n^2 = 16 \cdot n^2.$$

(Ω part next) To establish a lower bound we observe that since n is non-negative,

$$(n+3)^2 = n^2 + 6n + 9 \geq 1 \cdot n^2$$

Therefore for $c_1 = 1, c_2 = 16$ and for all $n \geq 1 = n_0$ we have that $(n+3)^2 = \Theta(n^2)$.
(c) Left as an exercise. ■

Exercise 448.
Show that $1000 = O(1)$.

Solution.
There exists constant $c = 2000$ such that $1000 \leq c \cdot 1 = 2000 \cdot 1 = 2000$, for all $n \geq 1 = n_0$. $c = 1000$ works also.
Be reminded: We don't need to find the best c or n_0. ■

Exercise 449.
TRUE or FALSE?

1. $\lg(n!) = O(n^2)$.
2. $n + \sqrt{n} = O(n^2)$.
3. $n^2 + \sqrt{n} = O(n^2)$.
4. $n^3 + 2\sqrt{n} = O(n^2)$.
5. $1/n^3 = O(\lg n)$.
6. $n^2\ sin^2(n) = \Theta(n^2)$. ($sin$ is the well-known trigonometric function).

Solution.
Omitted. ■

Exercise 450.
Prove the following.

1. $(n-10)^2 = \Theta(n^2)$.
2. $n^4 + 10n^3 + 100n^2 + 1890n + 98000 = \Omega(n^4)$.
3. $n^4 + 10n^3 + 100n^2 + 1890n + 98000 = \Omega(n^2)$.
4. $n^4 - 10n^3 - 100n^2 - 1890n + 100000 = O(n^4)$.
5. $n^2 - 20n - 20 = \Omega(n)$.
6. $n^2 + 20n = O(n^2)$.

Solution.
You prove. ■

Exercise 451.
Which of the two is asympotically larger? $f(n) = 2^{n/3}$ or $g(n) = 2^{(n/2)}$? Explain.

Solution.
Take the limit $\lim_{n\to\infty} f(n)/g(n)$.

$$\lim_{n\to\infty} f(n)/g(n) = 2^{n/3}/2^{n/2} = 0$$

Since it is zero $f(n)$ is asymptotically smaller than $g(n)$ and therefore $g(n)$ is asymptotically larger than $f(n)$. ■

Exercise 452.
What is the order of growth of

$$f(n) = 1 + x + x^2 + \ldots + x^n$$

for constant x, and integer n ($n \geq 1$) for the following three cases? Asymptotics are for n going to ∞. Justify your answer(s) and provide simplest possible answer(s).

1. x is a constant greater than 1,
2. $x = 1$,
3. $0 < x < 1$.

Solution.
If $x > 1$, the sum or equivalently $f(n)$ is

$$f(n) = \sum_{i=0}^{n} x^i = (x^{n+1} - 1)/(x - 1) = \Theta(x^n).$$

■

Exercise 453.

noindent Replace ? with each one of $\omega, o, \Theta, O, \Omega$ that makes the relationship between the two sides true. If none is applicable write "none". If necessary, use the approximation $n! \approx (n/e)^n$, where $e = 2.71\ldots$. Assume that $c > 1$ is a constant. Justify your answers.

1. $\lg(n!) = ?\ (n^5)$
2. $2^n = ?\ (8^{n/3})$
3. $n^{\lg n} = ?\ (4^{\lg n * \lg n})$
4. $1 = ?\ (100)$
5. $n^4 - n^3 = ?\ (n^4)$
6. $4^n = ?\ (3^{2n+c})$

Solution.
1. $\lg n! = o(n^5)$ since by Sterling's approximation to the factorial $\lg(n!) = \lg(n/e)^n = n\lg n - n\lg e$, and thus

$$\lim \frac{\lg n!}{n^5} = \lim \frac{n\lg n}{n^5} = \lim \frac{\lg n}{n^4} = 0$$

if in the latter, we apply L'Hopital's rule once at the end. If it is o, then it will also be O.
2. This is Θ and thus also O, Ω. Why? $8^{n/3} = (8^{1/3})^n = 2^n$.
3. $n^{\lg n} = 2^{\lg n * \lg n}$. This is because $2^{\lg n * \lg n} = \left(2^{\lg n}\right)^{\lg n} = n^{\lg n}$. Also $4^{\lg n * \lg n} = 2^{2\lg n \lg n} = 2^{\lg n \lg n} \cdot 2^{\lg n \lg n} = n^{\lg n} \cdot n^{\lg n}$. Therefore $\lim n^{\lg n}/4^{\lg n * \lg n} = \lim 1/n^{\lg n} = 0$. It is o and also O.
4. $1 = \Theta(100)$ and thus O, Ω as well.
5. Same arguments and answers as 4.
6. $3^{2n+c} = 9^n \cdot 3^c$. Taking limits $\lim(4^n/9^n) \cdot 1/3^c = 0$ since $(4/9)^n \to 0$, and $3^c \to 3^c$ since c is a constant, as $n \to \infty$. ■

Exercise 454.

The Fast vs the (furiously) Good. For some problem P we have implemented two algorithms that solve P. An algorithm `Good` runs in time $g(n)$ and an algorithm `Fast` runs in time $f(n)$, where n is problem size of P. Answer the following question Q for each one of the following three cases C1, C2 and C3. Justify your answer.

Q. Is *Fast* faster than *Good* for all $n \geq n_0$, for some constant n_0?

C1. $g(n) = \Omega(f(n) \lg n)$.

C2. $g(n) = \Theta(f(n) \lg n)$.

C3. $g(n) = O(f(n) \lg n)$.

Solution.

Asymptotic notation roundup: Think! ■

Exercise 455.
True or False? Explain.
f, g are asymptotically positive functions.
(a) If $f(n) = O(g(n)$ then $g(n) = O(f(n)$.
(b) If $f(n) + g(n) = O(\min(f(n), g(n)$.
(c) If $f(n) = O(g(n)$ then $2^{f(n)} = O(2^{g(n)})$.
(d) $f(n) = O(f(n)^2)$.
(e) $f(n) = \Theta(f(n/2))$.

Solution.
Use $f(n) = 1$ and $g(n) = n$ for the first two. Use $f(n) = 2\lg n$ and $g(n) = \lg n$ for the third. Use $f(n) = 1/n$ for the fourth. Use $f(n) = 2^n$ for the fifth. ■

7.4 Recurrence relation solution

Exercise 456.

Give a **TRUE** or **FALSE** for each one of the statements below. Any answer other than a full TRUE or FALSE will be considered wrong.
(a) The asymptotic solution of $T(n) = 16T(n/2) + n^3$ is $T(n) = \Theta(n^3)$.
(b) The asymptotic solution of $T(n) = 8T(n/2) + n^3 \lg n$ is $T(n) = \Theta(n^3 \lg n)$.
(c) The asymptotic solution of $T(n) = 2T(n/2) + \lg n$ is $T(n) = \Theta(n \lg n)$
(d) Let $f(n) > 0$ be a positive function for all $n \geq 0$. The following is true for every f: $f(n) = \Theta(f(n/2))$.
(e) The solution of $T(n) = T(n-1) + 1$ with $T(1)$ constant is $T(n) = \Theta(n^2)$.
(f) The solution of $T(n) = T(n-1) + n$ with $T(1)$ constant is $T(n) = \Theta(n^2)$.
(g) The asymptotic solution of $T(n) = T(n/2) + \lg n$ is $T(n) = \Theta(n)$.
(h) The asymptotic solution of $T(n) = 3T(n/2) + T(2n/4) + n^2$ is $T(n) = \Theta(n^2)$.
(i) The asymptotic solution of $T(n) = 32T(n/2) + n^4$, is $T(n) = \Theta(n^4)$.
(j) The asymptotic solution of $T(n) = T(n/2) + n$, is $T(n) = \Theta(n)$.
(k) The asymptotic solution of $T(n) = 4T(n/2) + n \lg n$ is $T(n) = \Theta(n(\lg n)^2)$.
(l) The asymptotic solution of $T(n) = 8T(n/8) + n$, is $T(n) = \Theta(n)$.
(m) The asymptotic solution of $T(n) = T(n/8) + n$, is $T(n) = \Theta(n)$.
(n) The asymptotic solution of $T(n) = 8T(n/2) + n^2$, is $T(n) = \Theta(n^2)$.
(o) The asymptotic solution of $T(n) = T(n/2) + n$, is $T(n) = \Theta(n)$.
(p) The asymptotic solution of $T(n) = 4T(n/2) + n^2$, is $T(n) = \Theta(n^2)$.
(q) The asymptotic solution of $T(n) = 16T(n/2) + n^5$, is $T(n) = \Theta(n^4)$.
(r) The asymptotic solution of $T(n) = 2T(n/4) + \Theta((n-1)!)$ is $T(n) = \Theta(n!)$. (factorial involved).
(s) The asymptotic solution of $T(n) = T(n/5) + T(7n/10) + n$ is $T(n) = \Theta(n)$.
(t) The asymptotic solution of $T(n) = 32T(n/2) + n^4$, is $T(n) = \Theta(n^4)$.
(u) The asymptotic solution of $T(n) = 16T(n/4) + n$ is $T(n) = \Theta(n^2)$.

Solution.
FALSE,FALSE,FALSE, FALSE
FALSE it is $\Theta(n)$, TRUE since $T(n) = T(n-1) + n = T(n-2) + (n-1) + n = \ldots = T(1) + 2 + 3 + \ldots + (n-1) + n$, FALSE; Case 2 of the master method shows that it is $T(n) = \Theta(\lg^2 n)$,
FALSE; Case 2 of the master method show that it is $T(n) = \Theta(n^2 \lg n)$. Note also that $3T(n/2) + T(2n/4) = 4T(n/2)$, FALSE; Case 1 of the master method case shows that it is $T(n) = \Theta(n^5)$, TRUE; Case 3 of the master method case,
FALSE; Case 1 implies $T(n) = \Theta(n^2)$, FALSE; Case 2 of the master method gives $T(n) = \Theta(n \lg n)$, TRUE; Case 3 of the master method gives $T(n) = \Theta(n)$, FALSE; Case 1 of the master method gives $T(n) = \Theta(n^3)$,
TRUE; Case 3 of the master method, FALSE; Case 2 of the master method gives $T(n) = \Theta(n^2 \lg n)$, FALSE; Case 3 of the master method gives $T(n) = \Theta(n^5)$, Think and do it, TRUE, FALSE, TRUE. ■

Exercise 457.

Solve the following recurrences. You may assume $T(1) = \Theta(1)$, where necessary. Make your bounds as tight as possible. Use asymptotic notation to express your answers. Justify your answers.

a. $T(n) = 4T(n/2) + 3n^2$
b. $T(n) = 32T(n/4) + 3n^2$
c. $T(n) = 4T(n/2) + n^3$
d. $T(n) = 9T(n/3) + n.$
e. $T(n) = 25T(n/5) + n^2.$

Solution.
One can use the master method for all of them.

a. $T(n) = 4T(n/2) + 3n^2$. It has $a = 4 \geq 1$, $b = 2 > 1$, $f(n) = 3n^2$. $n^{\log_b a} = n^{\lg 4} = n^2$. Obviously $n^2 = f(n) = \Theta(n^{\log_b a})$. Case 2 of the Master method applies and thus $T(n) = \Theta(n^2 \lg n)$.

b. $T(n) = 32T(n/4) + 3n^2$. Note that $32 \cdot 32 = 1024 = 2^{10}$, $4^5 = 1024$ i.e. $4^{5/2} = 1024^{1/2} = 32$. It has $a = 32 \geq 1$, $b = 4 > 1$, $f(n) = 3n^2$. $n^{\log_b a} = n^{\lg_4 32} = n^{5/2}$. Obviously $n^2 = f(n) = O(n^{\log_b a - \varepsilon}) = O(n^2)$, by choosing $\varepsilon = 1/2 > 0$. Case 1 of the Master method applies and thus $T(n) = \Theta(n^{5/2})$.

c. $T(n) = 4T(n/2) + n^3$. It has $a = 4 \geq 1$, $b = 2 > 1$, $f(n) = n^3$. $n^{\log_b a} = n^{\lg 4} = n^2$. Obviously $n^3 = f(n) = \Omega(n^{\log_b a + \varepsilon}) = O(n^3)$, by choosing $\varepsilon = 1 > 0$. Case 3 of the Master method applies since also $af(n/b) = 4(n/2)^3 = n^3/2 \leq cn^3 = (1/2)n^3$, by choosing $c = 1/2 < 1$, and thus $T(n) = \Theta(n^3)$.

d. We use the master method. $a = 9 \geq 1$, $b = 3 > 1$, $f(n) = n$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n = O(n^{\lg_b a - 1}) = O(n)$, case one of the master method applies, and thus $T(n) = \Theta(n^2)$.

e. We use the master method. $a = 25 \geq 1$, $b = 5 > 1$, $f(n) = n^2$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n^2 = \Theta(n^{\lg_b a}) = \Theta(n^2)$, case two of the master method applies, and thus $T(n) = \Theta(n^2 \lg n)$. ■

Exercise 458.

Solve the recurrence $T(n) \leq T(\lfloor n/2 \rfloor) + 1$ for $n \geq 2$, and $T(1) = 1$.
In order to avoid complications we will simplify things by disregarding issues with floors and ceilings: for example assum n is a power of two or more likely n is a power of two minus one or plus one!

Solution.

$$\begin{aligned} T(n) &= T(n/2) + 1 \\ &= (T(n/4) + 1) + 1 \\ &= T(n/2^2)2 * 1 \\ &= (T(n/2^3) +)2 * 1 \\ &= T(n/2^3) + 3 * 1 \\ &\cdots \\ &= T(n/2^i) + i * 1 \\ T(n) &= T(n/2^i) + i. \end{aligned}$$

The undolding of the chains ends when $T(n/2^i) = T(1)$ i.e. for $n/2^i = 1$. Solving for i we have $i = \lg n$. Thus

$$\begin{aligned} T(n) &= T(n/2^i) + i. \\ &= T(n/2^{\lg n}) + \lg n \\ &= T(1) + \lg n \\ &= \lg n + 1. \end{aligned}$$

In fact $T(n)$ is the maximum (or worst-case) running number of comparisons, since it is possible that we found our key with the first comparison performed! Thus $T(n) \leq \lg n + 1$ might be a better (and correct) answer. If we re-consider the fact that n is a power of n and thus $n/2$ does not include the floor function any more, we need to utilize asymptotic notation and say $T(n) = O(\lg n)$. In such a notation the plus one disappears as the most important term is the logarithmic term. There is an added benefit in using this form. We indeed describe not just the number of comparisons but also the running time now. Everything else in Binary Search revolves around a comparison: we increment a variable i or l, decrement a variable r or j, or "compare" variables i and j or l and r plus the $(r+l)/2$ or $(i+j)/2$ or its variants. Thus all auxiliary operations are proportional to $\lg n$ i.e. linear to $\lg n$ also captured in the form $T(n) = O(\lg n)$.
And food for through is $i < j$ or $i \leq j$ a COMPARISON as used earlier? The answer is NO. i and j are integer values not keys. Functions cmp expects objects that are keys. We compare i and j using a subtraction $i - j$ involving the accumulator of a CPU. If the result is zero i.e. $i = j$ the FLAGS set this immediately. If the accumulator is positive $i > j$ or negative $i < j$ this become immediately available in one instruction that is a subtraction i.e nanoseconds. Comparing keys or key values takes a function call i.e. a context switch or microseonds. (And if you argue otherwise keys can be quite long, required multiple reads from main memory and a read is 80-100ns each.) ■

Exercise 459.
Solve the recurrence relation using the master method.

$$T(n) = 2T(n/2) + n.$$

Solution.
$a = 2 > 0$,
$b = 2 > 1$ and
$\lg_b a = 1$ i.e. $n^{\lg_b a} = n$,
$f(n) = n$

Obviously $f(n) = \Theta(n^{\lg_b a})$ i.e. $n = \Theta(n)$ thus **CASE 2** of the Master method is applicable. Then $T(n) = \Theta(n \lg n)$. ■

Exercise 460.
Solve the recurrence relation using the master method.

$$T(n) = 8T(n/2) + n^2.$$

Solution.
$a = 8 > 0$,
$b = 2 > 1$ and
$\lg_b a = \lg_2 8 = 3$ i.e. $n^{\lg_b a} = n^3$,
$f(n) = n^2$

Obviously $f(n) = O(n^{\lg_b a - 1})$ i.e. $n^2 = \Theta(n^{3-1})$ thus **CASE 1** is applicable with $\varepsilon = 1 > 0$. Then $T(n) = \Theta(n^3)$. ■

Exercise 461.
Solve the recurrence relation using the master method.

$$T(n) = 2T(n/2) + n^2.$$

Solution.
$a = 2 > 0$,
$b = 2 > 1$ and
$\lg_b a = 1$ i.e. $n^{\lg_b a} = n$,
$f(n) = n^2$

Obviously $f(n) = \Omega(n^{\lg_b a + 1})$ i.e. $n^2 = \Omega(n^{1+1})$ with $\varepsilon = 1 > 0$, i.e. **CASE 3** might be applicable. For that we need to verify the second condition $af(n/b) \leq cf(n)$.

$$\begin{aligned} af(n/b) &\leq cf(n) \\ 2f(n/2) &\leq cf(n) \\ 2(n/2)^2 &\leq cn^2 \\ n^2/2 &\leq cn^2 \end{aligned}$$

If we choose $c = 1/2 < 1$ the secondary condition is also true. Then $T(n) = \Theta(n^2)$. ■

Exercise 462.
$T(n) = T(n/2) + f(n)$, where $a = 1$ and $b = 2$, and $n^{\lg_b a} = n^0 = 1$.

Solution.
If $f(n) = 1$, then $T(n) = \Theta(\lg n)$.
If $f(n) = \sqrt{n}$, then because $af(n/b) = a\sqrt{n/b} = 1\sqrt{n}/\sqrt{2} \leq \sqrt{1/2}\sqrt{n} = cf(n)$ for $c = \sqrt{1/2}$, we get that $T(n) = \Theta(\sqrt{n})$. ■

Exercise 463.

The iteration method is also know as the recursion tree method or the repeated substitution method.
Let n be a power of two. Solve exactly the following recurrence.
$T(n) = T(n/2) + n$, where $T(1) = 2$.

Solution.
We could use the master method (case three) to derive that $T(n) = \Theta(n)$ (using $\varepsilon = 1/2$ and $c = 1/2$), if boundary condition $T(1) = 2$ was not explicitly stated. Since it is, we use the recursion tree/iteration method, in fact the iteration method.

$$\begin{aligned}
T(n) &= T(n/2) + n \\
&= T(n/4) + n/2 + n \\
&= T(n/8) + n/4 + n/2 + n \\
&= T(n/2^i) + n/2^{i-1} + \ldots + n/4 + n/2 + n \\
&= T(n/2^{\lg n}) + n/(n/2) + \ldots + n/2 + n \\
&= T(1) + 2 + 4 + 8 + \ldots + n/2 + n \\
&= 2 + 2 + 4 + 8 + \ldots + n/2 + n \\
&= 1 + (1 + 2 + 4 + 8 + \ldots + n/2 + n) \\
&= 1 + (2n - 1) = 2n
\end{aligned}$$

We used the fact that $1 + 2 + 4 + \ldots + n = 2n - 1$. ■

Exercise 464.
Solve the recurrence

$$T(n) = T(n-1) + 2n, \quad \text{where} \quad T(1) = 2.$$

Solution.
We use the iteration method.

$$\begin{aligned}
T(n) &= T(n-1) + 2n = (T(n-2) + 2(n-1)) + 2n \\
&= T(n-2) + 2(n-1) + 2n = (T(n-3) + 2(n-2)) + 2(n-1) + 2n \\
&= T(n-3) + 2(n-2) + 2(n-1) + 2n \\
&= \dots \\
&= T(i) + 2(i+1) + \dots + 2(n-1) + 2n \\
&= T(1) + 2(1+1) + \dots + 2(n-1) + 2n \\
&= 2 \cdot 1 + 2(1+1) + \dots + 2(n-1) + 2n \\
&= 2(1 + 2 + \dots + n) \\
&= 2n(n+1)/2 = n(n+1)
\end{aligned}$$

■

Exercise 465.
Solve the following recurrence exactly. You may assume n is a power of two.

$$T(n) = 2T(n/2) + 6n,$$

where $T(4) = 12$.

Solution.

$$\begin{aligned} T(n) &= 2T(n/2) + 6n \\ &= 2(2T(n/2^2) + 6n/2) + 6n = 2^2 T(n/2^2) + 2(6n) \\ &= 2^3 T(n/2^3) + 3(6n) \\ &\cdots \\ &= 2^i T(n/2^i) + i(6n) \end{aligned}$$

Substituting $n/2^i = 4$ and solving for i we get $i = \lg n - 2$. Then $T(n/2^i) = T(4)$ and $T(4) = 12$. Substituting $\lg n - 2$ for i in.

$$\begin{aligned} T(n) &= 2^i T(n/2^i) + i(6n) \\ &= (n/4)T(4) + (\lg n - 2)6n \\ &= 12n/4 + 6n\lg n - 12n \\ &= 6n\lg n - 9n \end{aligned}$$

The result follows. ■

Exercise 466.
Solve exactly the following recurrence. You may assume that n is a power of two.

$$T(n) = 2T(n/2)+18n, \qquad T(8)=6.$$

Do the same for

$$T(n) = 2T(n/2)+6n, \qquad T(8)=18.$$

Solution.

$$\begin{aligned}
T(n) &= 2T(n/2)+18n, \qquad T(8)=6.\\
&= 2^2T(n/2^2)+2\cdot 18n\\
&= 2^iT(n/2^i)+i\cdot 18n\\
&= 2^{\lg n-3}T(n/2^{\lg n-3})+(\lg n-3)\cdot 18n\\
&= (n/8)T(8)+18n\lg n-54n\\
&= (n/8)6+18n\lg n-54n\\
&= 18n\lg n-53.25n
\end{aligned}$$

We used in the fourth equality the facts that $T(8)=6$ and $2^{\lg n-3}=n/8$.

$$\begin{aligned}
T(n) &= 2T(n/2)+6n, \qquad T(8)=18.\\
&= 2^2T(n/2^2)+2\cdot 6n\\
&= 2^iT(n/2^i)+i\cdot 6n\\
&= 2^{\lg n-3}T(n/2^{\lg n-3})+(\lg n-3)\cdot 6n\\
&= (n/8)T(8)+6n\lg n-18n\\
&= (n/8)18+6n\lg n-18n\\
&= 6n\lg n-15.75n
\end{aligned}$$

We used in the fourth equality the facts that $T(8)=18$ and $2^{\lg n-3}=n/8$. ■

Exercise 467.

Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 3.

$$T(n) = 3T(n/3) + \lg_3 n, \text{ where } T(1) = 1.$$

Solution.
We have that

$$\sum_{k=0}^{m-1} kx^k = \frac{mx^m(x-1) - x^{m+1} + x}{(x-1)^2} \; and \; for \; x = 3 \; \rightarrow \sum_{k=0}^{m-1} k3^k = (2m3^m - 3^{m+1} + 3)/4$$

Then,

$$\begin{aligned} T(n) &= 3T(n/3) + \lg_3 n \\ &= 3^{\lg_3 n} T(1) + 3^{\lg_3 n - 1} \lg_3 \frac{n}{3^{\lg_3 n - 1}} + \ldots + 3^i \lg_3 \frac{n}{3^i} + \ldots + 3^1 \lg_3 \frac{n}{3^1} + 3^0 \lg_3 \frac{n}{3^0}. \end{aligned}$$

Therefore, by substituting for $T(1) = 1$ we get

$$T(n) = 3^{\lg_3 n} + \lg_3 n(1 + 3 + \ldots + 3^{\lg_3 n - 1}) - \sum_{k=0}^{\lg_3 n - 1} k3^k = n + \lg_3 n(n-1)/2 - \lg_3 n \, n/2 + 3n/4 - 3/4 = 7n/4 - \lg_3 n/2 - 3/4.$$

■

Exercise 468.

Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 4, i.e. $n = 4^k$.

$$T(n) = 4T(n/4) + n^2, \text{ where } T(1) = 4.$$

Solution.
$T(n) = 4T(n/4) + n^2$. We use the iteration method. **This first example will be done in fine detail showing all calculations**.

$$\begin{aligned}
T(n) &= 4T(n/4) + n^2 \\
&= 4\left(4T\left(\frac{n}{4^2}\right) + \left(\frac{n}{4}\right)^2\right) + n^2 \\
&= 4^2 T\left(\frac{n}{4^2}\right) + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^2\left(4T\left(\frac{n}{4^3}\right) + \left(\frac{n}{4^2}\right)^2\right) + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^3 T\left(\frac{n}{4^3}\right) + \frac{n^2}{4^2} + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= \dots \\
&= 4^i T\left(\frac{n}{4^i}\right) + \frac{n^2}{4^{i-1}} + \dots + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{1}{4^{i-1}} + \dots + \frac{1}{4^1} + \frac{1}{4^0}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{\frac{1}{4^i} - 1}{\frac{1}{4} - 1}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{1 - \frac{1}{4^i}}{1 - \frac{1}{4}}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + \frac{4n^2}{3}\left(1 - \frac{1}{4^i}\right)
\end{aligned}$$

The base case is $T(1) = 4$. We set $n/4^i = 1$, i.e. $n = 4^i$. If we solve this for i, we get $i = \log_4 n = \lg n/2$. Then $T(n/4^i) = T(1) = 4$. In addition $4^i = n$. Therfore the formula for $T(n)$ becomes.

$$\begin{aligned}
T(n) &= 4^i T\left(\frac{n}{4^i}\right) + \frac{4n^2}{3}\left(1 - \frac{1}{4^i}\right) \\
&= nT(1) + \frac{4n^2}{3}(1 - 1/n) \\
&= 4n + \frac{4n^2}{3}(1 - 1/n) \\
&= \frac{4n^2}{3} + \frac{8n}{3}
\end{aligned}$$

Note. It is easy to verify for example that $T(1) = \frac{4\cdot 1^2}{3} + \frac{8\cdot 1}{3} = 4$. This can serve as a checking mechanism that the solution that you have established satisfies at least the base case. ■

Exercise 469.

Solve exactly using the recursion tree method or the iteration method the following recurrence. You may assume that n is a power of 2.

$$T(n) = 4T(n/2) + n$$

with $T(4) = 6$.

Solution.

$$\begin{aligned}
T(n) &= 4T(n/2) + n \\
&= 4\left(4T\left(\frac{n}{2^2}\right) + \left(\frac{n}{2}\right)\right) + n \\
&= 4^2 T\left(\frac{n}{2^2}\right) + \frac{4^1 n}{2^1} + n \\
&= 4^2\left(4T\left(\frac{n}{2^3}\right) + \left(\frac{n}{2^2}\right)\right) + 2^1 n + 2^0 n \\
&= 4^3 T\left(\frac{n}{2^3}\right) + 2^2 n + 2^1 n + 2^0 n \\
&= \dots \\
&= 4^i T\left(\frac{n}{2^i}\right) + 2^{i-1} n + \ldots + 2^1 n + 2^0 n \\
&= 4^i T\left(\frac{n}{2^i}\right) + n\left(2^i - 1\right)
\end{aligned}$$

The base case is $T(4) = 6$. We set $n/2^i = 4$, i.e. $n = 2^{i+2}$. If we solve this for i, we get $i = \lg n - 2$. Then $T(n/2^i) = T(4) = 6$. In addition, $2^i = n/4$, $4^i = 2^i \cdot 2^i = n/4 \cdot n/4 = n^2/16$. Therfore the formula for $T(n)$ becomes.

$$\begin{aligned}
T(n) &= 4^i T\left(\frac{n}{2^i}\right) + n\left(2^i - 1\right) \\
&= (n^2/16)T(4) + n(n/4 - 1) \\
&= 3n^2/8 + n^2/4 - n \\
&= 5n^2/8 - n
\end{aligned}$$

Note. It is easy to verify for example that $T(4) = 5 \cdot 4^2/8 - 4 = 6$. This can serve as a checking mechanism that the solution that you have established satisfies at least the base case. ■

Exercise 470.

Solve exactly using the recursion tree method or the iteration method the following recurrence. You may assume that n is a power of 2.

$$T(n) = 8T(n/4) + n$$

with $T(1) = 1$.

Solution.

$$\begin{aligned} T(n) &= 8T(n/4) + n \\ &= 8(8T(n/4^2) + n/4) + n \\ &= 8^2 T(n/4^2) + 2^1 n + 2^0 n \\ &= 8^3 T(n/4^3) + 2^2 n + 2^1 n + 2^0 n \\ &= 8^i T(n/4^i) + 2^{i-1} n + \ldots + 2^1 n + 2^0 n \\ &= 8^i T(n/4^i) + (2^i - 1) \cdot n \end{aligned}$$

For $n/4^i = 1$ we have $i = \lg n/2$ and thus $T(n/4^i) = T(1) = 1$. Then the recurrence gives.

$$\begin{aligned} T(n) &= 8^i T(n/4^i) + (2^i - 1) \cdot n \\ &= 8^{\lg n/2} \cdot 1 + (2^{\lg n/2} - 1)n \\ &= 8^{\lg n/2} \cdot 1 + (n^{1/2} - 1)n \\ &= n^{3/2} + n^{3/2} - n \\ &= 2n^{3/2} - n \end{aligned}$$

Therefore, $T(n) = 2n^{3/2} - n$.

■

Exercise 471.

Solve exactly using the recursion tree method or the iteration method the following recurrence. You may assume that n is a power of 2.

$$T(n) = 8T(n/2) + n^2, \text{ where } T(1) = 1.$$

Solution.

$$\begin{aligned} T(n) &= 8T(n/2) + n^2 \\ &= 8(8T(n/2^2) + n^2/2^2) + n^2 \\ &= 8^2T(n/2^2) + 2^1n^2 + n^2 \\ &= 8^3T(n/2^3) + 2^2n^2 + 2^1n^2 + n^2 \\ &= 8^iT(n/2^i) + 2^{i-1}n^2 + \ldots + 2^1n^2 + n^2 \\ &= 8^iT(n/2^i) + (2^i - 1)\cdot n^2 \end{aligned}$$

For $n/2^i = 1$ we have $i = \lg n$ and thus $T(n/2^i) = T(1) = 1$. Then the recurrence gives.

$$\begin{aligned} T(n) &= 8^iT(n/2^i) + (2^i - 1)\cdot n^2 \\ &= 8^{\lg n}\cdot 1 + (n-1)n^2 \\ &= n^3 + (n-1)n^2 \\ &= 2n^3 - n^2 \end{aligned}$$

Therefore, $T(n) = 2n^3 - n^2$.

■

Exercise 472.

Solve exactly using the recursion tree method the following recurrence. You may assume that n is a power of 3.

$$T(n) = 9T(n/3) + n^2, \text{ where } T(1) = 1.$$

Solution.

$$\begin{aligned} T(n) &= 9T(n/3) + n^2 \\ &= 9(9T(n/3^2) + n^2/9) + n^2 \\ &= 9^2 T(n/3^2) + n^2 + n^2 \\ &= 9^i T(n/3^i) + n^2 + \ldots + n^2 + n^2 \\ &= 9^i T(n/3^i) + i \cdot n^2 \end{aligned}$$

A first observation is that the number of n^2's we have is equal to the power of 9 (or equivalently, of 3) in the coefficient of $T(.)$ (or the denominator of $T(n/.)$ respectively). In the base case $T(1) = 1$, therefore we choose $n/3^i = 1$, which gives $i = \lg n / \lg 3$. Then, for this i, $T(n/3^i) = T(1) = 1$. We note that $9 = 3^2 = (2^{\lg 3})^2 = 2^{2\lg 3}$. We therefore get that

$$\begin{aligned} T(n) &= 9^i T(n/3^i) + in^2 \\ &= 9^{\lg n/\lg 3} T(1) + n^2 \lg n/\lg 3 \\ &= 3^{2\lg n/\lg 3} + n^2 \lg n/\lg 3 \\ &= 2^{2\lg 3 \lg n/\lg 3} + n^2 \lg n/\lg 3 \\ &= 2^{2\lg n} + n^2 \lg n/\lg 3 \\ &= n^2 + n^2 \lg n/\lg 3 \end{aligned}$$

It is easy to verify that this solution satisfies both the recurrence and the boundary condition. ■

Exercise 473.

Solve the following recurrences. Make your bounds as tight as possible. You may assume that n is a power of two.

$$T(n) = T(n/2) + \lg n, \qquad \text{where} \quad T(2) = 1.$$

Solution.
$T(n) = T(n/2) + \lg n$. We use the iteration method, noting that $T(2) = 1$.

$$\begin{aligned} T(n) &= T\left(\frac{n}{2}\right) + \lg n \\ &= \left(T\left(\frac{n}{2^2}\right) + \left(\lg \frac{n}{2}\right)\right) + \lg n \\ &= T\left(\frac{n}{2^2}\right) + \lg \frac{n}{2^1} + \lg \frac{n}{2^0} \\ &= \cdots \\ &= T\left(\frac{n}{2^i}\right) + \lg \frac{n}{2^{i-1}} + \ldots + \lg \frac{n}{2^1} + \lg \frac{n}{2^0} \end{aligned}$$

The base case is $T(2) = 1$. We set $n/2^i = 2$, i.e. $n = 2^{i+1}$. If we solve this for i, we get $i = \lg n - 1$. Then $T(n/2^i) = T(2) = 1$. In addition $2^i = n/2$. Therfore the formula for $T(n)$ becomes.

$$\begin{aligned} T(n) &= T\left(\frac{n}{2^i}\right) + \lg \frac{n}{2^{i-1}} + \ldots + \lg \frac{n}{2^1} + \lg \frac{n}{2^0} \\ &= T\left(\frac{n}{2^{\lg n - 1}}\right) + \lg \frac{n}{2^{\lg n - 1 - 1}} + \ldots + \lg \frac{n}{2^1} + \lg \frac{n}{2^0} \\ &= T(2) + \left(\lg n - (\lg 2^{\lg n - 1 - 1})\right) + \left(\lg n - (\lg 2^{\lg n - 2 - 1})\right) + \ldots + \left(\lg n - \lg 2^1\right) + \left(\lg n - \lg 2^0\right) \\ &= 1 + (\lg n - (\lg n - 2)) + (\lg n - (\lg n - 3)) + \ldots + (\lg n - 1) + (\lg n - 0) \\ &= 1 + 2 + 3 + \ldots + (\lg n - 1) + (\lg n) \\ &= \sum_{j=1}^{\lg n} j \\ &= \frac{\lg n (\lg n + 1)}{2} \end{aligned}$$

Therefore, $T(n) = \lg n(\lg n + 1)/2$. It is easy to verify that $T(2)$ is indeed 1. ■

Exercise 474.
Solve exactly, assuming n is a power of two, $T(n) = 2T(n/2) + \lg n$ where $T(1) = 1$.

Solution.

$$T(n) = 2T(n/2) + \lg n = \ldots = 2^{\lg n} T(1) + 2^{\lg n - 1} \lg \frac{n}{2^{\lg n - 1}} + \ldots + 2^i \lg \frac{n}{2^i} + \ldots + 2^1 \lg \frac{n}{2^1} + 2^0 \lg \frac{n}{2^0}.$$

Therefore, by substituting for $T(1) = 1$ we get

$$T(n) = 2^{\lg n} + \lg n(1 + 2 + \ldots + 2^{\lg n - 1}) - \sum_{k=0}^{\lg n - 1} k 2^k = n + (n-1) \cdot \lg n - n \cdot \lg nn + 2n - 2 = 3n - \lg n - 2.$$

■

Exercise 475.

The recurrence to solve is the following one for n a power of two.

$$T(n) = 2T(n/2) + (n \lg n), \text{ where } T(1) = 1.$$

Solution.
We solve the recurrence using the iteration method. Below, we set $n/2^i = 1$ to satisfy the boundary condition $T(1)$. Then, $i = \lg n$.

$$\begin{aligned}
T(n) = T(2^k) &= 2T(n/2) + (n \lg n) \\
&= 2^2 T(n/4) + 2(n/2)\lg(n/2) + n \lg n \\
&= 2^2 T(n/4) + n \lg(n/2) + n \lg n \\
&= 2^i T(n/2^i) + n \lg(n/2^{i-1}) + \ldots + n \lg(n/2^0) \\
&= 2^i T(n/2^i) + n \sum_{k=0}^{i-1} \lg(n/2^k) = 2^{\lg n} T(1) + n \sum_{k=0}^{\lg n - 1} (\lg(n) - k \lg 2) \\
&= n + n \sum_{k=0}^{\lg n - 1} \lg n - n \sum_{k=0}^{\lg n - 1} k \\
&= n + n \lg^2 n - n(\lg n - 1)(\lg n)/2 \\
&= n + n \lg n (\lg n + 1)/2
\end{aligned}$$

Therefore, $T(n) = \Theta(n \lg^2 n)$.

■

Exercise 476.

Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 3, ie $n = 3^k$.

$$T(n) = 3T(n/3) + 9n, \text{ where } T(1) = 1.$$

Solution.
Let $n = 3^k$ and thus $k = \log_3 n$. As $T(n) = 3T(n/3) + 9n$, then $T(n/3) = 3T(n/3) + 9n/3$ and similarly for $T(n/3^2) = \ldots$ and so on.

$$\begin{aligned}
T(n) &= 3\,T(n/3) + 9n \\
T(3^k) &= 3\,T(3^{k-1}) + 9\,3^k \\
&= 3\,(3\,T(3^{k-2}) + 9\,3^{k-1}) + 9\,3^k \\
&= 3^2\,T(3^{k-2}) + 9\,3^k + 9\,3^k \\
&= 3^2\,(3\,T(3^{k-3}) + 9\,3^{k-2}) + 9\,3^k + 9\,3^k \\
&= 3^3\,T(3^{k-3}) + 9\,3^k + 9\,3^{k-1} + 9\,3^k \\
&\ldots \\
&= 3^k\,T(3^{k-k}) + 9\,3^k + \ldots + 9\,3^k + 9\,3^k \\
&= 3^k\,T(1) + 9(3^k + \ldots + 3^{k-1} + 3^k) \\
&= 3^k\,1 + 9\,3^k\,k \\
&= n + 9n\log_3 n
\end{aligned}$$

■

Exercise 477.
Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 5.

$$T(n) = 5T(n/5) + n \lg_5 n, \text{ where } T(1) = 1.$$

Solution.
Let $n = 5^k$ and thus $k = \log_5 n$. As $T(n) = 5T(n/5) + n\lg_5 n$, then $T(5^k) = 5T(5^{k-1}) + k5^k$.

$$\begin{aligned}
T(n) &= 5\,T(n/5) + n\lg_5 n \\
T(5^k) &= 5\,T(5^{k-1}) + k5^k \\
&= 5\,(5\,T(5^{k-2}) + (k-1)5^{k-1}) + k5^k \\
&= 5^2\,T(5^{k-2}) + (k-1)5^k + k5^k \\
&= 5^2\,(5\,T(5^{k-3}) + (k-2)5^{k-2}) + (k-1)5^k + k5^k \\
&= 5^3\,T(5^{k-3}) + (k-2)5^k + (k-1)5^k + k5^k \\
&\dots \\
&= 5^k\,T(5^{k-k}) + 1\cdot 5^k + \ldots + (k-1)5^k + k5^k \\
&= 5^k\,T(1) + 5^k(1+2+\ldots+k) \\
&= 5^k\,1 + 5^k\,k(k+1)/2 \\
&= n + n\log_5(n)(\log_5(n)+1)/2
\end{aligned}$$

■

Exercise 478.

Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 2.

$$T(n) = 8T(n/2) + n, \text{ where } T(1) = 1.$$

Solution.
We are reminded of the geometric sequence $1 + x^1 + x^2 + \ldots + x^{i-1} = (x^i - 1)/(x-1)$. If $x = 4$ this gives $1 + 4^1 + 4^2 + \ldots + 4^{i-1} = (4^i - 1)/(4-1)$.

$$\begin{aligned} T(n) &= 8T(n/2) + n \\ &= 8(8T(n/2^2) + n/2) + n \\ &= 8^2 T(n/2^2) + 4^1 n + 4^0 n \\ &= 8^2(8T(n/2^3) + n/2^2) + 4^1 n + 4^0 n \\ &= 8^3 T(n/2^3) + 4^2 n + 4^1 n + 4^0 n \\ &= 8^i T(n/2^i) + 4^{i-1} n + \ldots + 4^1 n + 4^0 n \\ &= 8^i T(n/2^i) + ((4^i - 1)/(4-1)) \cdot n \end{aligned}$$

For $n/2^i = 1$ we have $i = \lg n$ and thus $T(n/2^i) = T(1) = 1$. We are also reminded that $8^{\lg n} = (2^3)^{\lg n} = 2^{3\lg n} = (2^{\lg n})^3 = n^3$. For the same reason, $4^{\lg n} = n^2$. Then the recurrence gives.

$$\begin{aligned} T(n) &= 8^i T(n/2^i) + (4^i - 1)n/3 \\ &= 8^{\lg n} + (4^{\lg n} - 1)n/3 \\ &= 8^{\lg n} + (n^2 - 1)n/3 \\ &= n^3 + (n^2 - n)/3 \\ &= 4n^3/3 - n/3 \end{aligned}$$

Therefore, $T(n) = 4n^3/3 - n/3$.

■

Exercise 479.

Solve exactly using the iteration method the following recurrence. You may assume that n is a power of 2.

$$T(n) = 4T(n/4) + n^2, \text{ where } T(1) = 4.$$

Solution.
We use the iteration method.

$$\begin{aligned}
T(n) &= 4T(n/4) + n^2 \\
&= 4\left(4T\left(\frac{n}{4^2}\right) + \left(\frac{n}{4}\right)^2\right) + n^2 \\
&= 4^2 T\left(\frac{n}{4^2}\right) + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^2\left(4T\left(\frac{n}{4^3}\right) + \left(\frac{n}{4^2}\right)^2\right) + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^3 T\left(\frac{n}{4^3}\right) + \frac{n^2}{4^2} + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= \dots \\
&= 4^i T\left(\frac{n}{4^i}\right) + \frac{n^2}{4^{i-1}} + \dots + \frac{n^2}{4^1} + \frac{n^2}{4^0} \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{1}{4^{i-1}} + \dots + \frac{1}{4^1} + \frac{1}{4^0}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{\frac{1}{4^i} - 1}{\frac{1}{4} - 1}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + n^2\left(\frac{1 - \frac{1}{4^i}}{1 - \frac{1}{4}}\right) \\
&= 4^i T\left(\frac{n}{4^i}\right) + \frac{4n^2}{3}\left(1 - \frac{1}{4^i}\right)
\end{aligned}$$

The base case is $T(1) = 4$. We set $n/4^i = 1$, i.e. $n = 4^i$. If we solve this for i, we get $i = \log_4 n = \lg n/2$. Then $T(n/4^i) = T(1) = 4$. In addition $4^i = n$. Therfore the formula for $T(n)$ becomes.

$$\begin{aligned}
T(n) &= 4^i T\left(\frac{n}{4^i}\right) + \frac{4n^2}{3}\left(1 - \frac{1}{4^i}\right) \\
&= nT(1) + \frac{4n^2}{3}(1 - 1/n) \\
&= 4n + \frac{4n^2}{3}(1 - 1/n) \\
&= \frac{4n^2}{3} + \frac{8n}{3}
\end{aligned}$$

Note. It is easy to verify for example that $T(1) = \frac{4\cdot 1^2}{3} + \frac{8\cdot 1}{3} = 4$. This can serve as a checking mechanism that the solution that you have established satisfies at least the base case. ■

Exercise 480.

Solve the following recurrence $T(n) = T(n/5) + T(7n/10) + n$ where $T(5) = 10$.

Solution.

Solution Preliminaries. This is a the recurrence for worst-case linear time selection (k order statistic problem) of a later Subject. Thus we know that solution is supposed to be linear i.e. $T(n) = \Theta(n)$. We prove the result in two steps first $T(n) = O(n)$ and then $T(n) = \Omega(n)$, utilizing the definition of O and Ω $T(n) = \Omega(n)$**.** We will find two positive constant c_1, n_0 such that $c_1 n \leq T(n)$ for all $n \geq n_0$. We define $\boxed{P(n) : T(n) \geq c_1 n}$.

Base case ($n = 5$). Show $P(5)$ is true. This is equivalent to $T(5) \geq c_1 \cdot 5$. For the latter to be true utilizing the base case we need to have the following.

$$\begin{array}{rcl} P(5) & is & \text{true?} \\ T(5) & \geq_? & c_1 \cdot 5 \\ 10 & \geq_? & c_1 \cdot 5 \\ 2 & \geq_? & c_1 \end{array}$$

Thus if $c_1 \leq 2$, then $P(5)$ is true and base case is proved.

Inductive Step.

$$P(5) \wedge \ldots \wedge P(n/5) \wedge \ldots \wedge P(7n/10) \wedge \ldots \wedge P(n-1) \Rightarrow P(n)$$

There is not much to be done. Starting with the recurrence $T(n) = T(n/5) + T(7n/10) + n \geq 0 + 0 + n$ we conclude that $T(n) \geq 1 \cdot n$.

Conclusion. From the base case we have $T(n) \geq c_1 n$ for $c_1 \leq 2$. From the inductive we have $T(n) \geq c_1 n$ with $c_1 \leq 1$. Reconciling the two we pick $c_1 = 1$ with n_0 the base case value i.e. $n_0 = 5$.

We continue with the second part and proving that $T(n) = O(n)$**.** We will find two positive constant c_2, n_0 such that $T(n) \leq c_2 n$ for all $n \geq n_0$. We define $\boxed{P(n) : T(n) \leq c_2 n}$.

Base case ($n = 5$). Show $P(5)$ is true. This is equivalent to $T(5) \leq c_2 \cdot 5$. For the latter to be true utilizing the base case we need to have the following.

$$\begin{array}{rcl} P(5) & is & \text{true?} \\ T(5) & \leq_? & c_2 \cdot 5 \\ 10 & \leq_? & c_2 \cdot 5 \\ 2 & \leq_? & c_2 \end{array}$$

Thus if $c_2 \geq 2$, then $P(5)$ is true and base case is proved.

Inductive Step.

$$P(5) \wedge \ldots \wedge P(n/5) \wedge \ldots \wedge P(7n/10) \wedge \ldots \wedge P(n-1) \Rightarrow P(n)$$

$\boxed{\textbf{X: } P(n/5)}$ **and** $\boxed{\textbf{Y : } P(7n/10)}$**.** The induction hypothesis implies $P(n/5)$ and $P(7n/10)$ i.e. $\boxed{\text{X: } T(n/5) \leq c_2 n/5}$ and $\boxed{\text{Y: } T(7n/10) \leq c_2(7n/10)}$. We then use the recurrence and X and Y as follows.

$$\begin{array}{rcl} T(n) & = & T(n/5) + T(7n/10) + n \\ & \leq_X & c_2 n/5 + T(7n/10) + n \\ & \leq_Y & c_2 n/5 + c_2(7n/10) + n \\ & \leq & c_2(9n/10) + n \end{array}$$

We have shown so far that $T(n) \leq c_2(9n/10)+n$. If we manage to prove that $c_2(9n/10)+n \leq c_2 n$ then by transitivity we have shown $P(n) : T(n) \leq c_2 n$ and we are done.

$$\begin{aligned} c_2(9n/10)+n &\leq_? c_2 n \\ n &\leq_? c_2 n - c_2(9n/10) \\ n &\leq_? c_2 n/10 \\ 1 &\leq_? c_2/10 \\ 10 &\leq_? c_2 \end{aligned}$$

In the previous derivation n was cancelled out provided $n > 0$. Since by the definition of big-Oh n is integer we use $n \geq 1$. Now if $T(n) \leq c_2(9n/10)+n$ and $c_2(9n/10) \leq c_2 n$ provided $n \geq 1$ and $c_2 \geq 10$ we have also shown by transitivity $T(n) \leq c_2 n$ and thus we conclude the inductive step.
Conclusion. By the base case $c_2 \geq 2$ and $n_0 = 5$. By the inductive step $n_0 = 1$ and $c_2 \geq 10$. To reconcile we use

$n_0 = 5$ and $c_2 = 10$.
O **and** Ω **reconciliation.** For the Ω part we used $n_0 = 5$ and $c_1 = 1$. For the O part we have $n_0 = 5$ and $c_2 = 10$.

Thus $c_1 = 1, c_2 = 10, n_0 = 5$ and

$$1 \cdot n \leq T(n) \leq 10 \cdot n \qquad \forall n \geq 5.$$

■

Exercise 481.

Solve the recurrence $T(n) = 2T(n/2) + n$ using the substitution (a.k.a. guess-and-check) method. (Implicit assumption is that $T(n)$ is nonnegative and defined for all positive n, or for arbitrarily large n).

Solution.
Since no boundary condition is given we can thus choose k and l constants greater than zero so that $T(k) = l$. We choose k, l in such a way to make the inductive proof as simple as possible. Let us choose $T(1) = 0$.
A. Guess Step: $T(n) = O(n \lg n)$. We guess $T(n) = O(n \lg n)$, i.e. $\exists$ pos. constant $c_2, n_0 : T(n) \leq c_2 n \lg n$, $\forall n \geq n_0$.
A.0 Check Step (Strong Induction). We shall prove our claim by using induction. Our $P(n)$, the proposition to be proven true would be $T(n) \leq c_2 n \lg n$ for arbitrarily large values of $n \geq n_0$, for some positive constant n_0.
A.1. Base Case of Induction. We show $P(1)$ is true is $T(1) \leq c_2 \cdot 1 \cdot \lg 1$. Since $T(1) = 0$ it is trivially true that $0 = T(1) \leq c_2 \cdot 1 \cdot \lg 1 = 0$ for all choices of a positive and constant c_2. Base case completed.
A.2. Inductive Step. $P(1) \wedge \ldots \wedge P(n/2) \wedge \ldots \wedge P(n-1) \Rightarrow P(n)$. We show that if $P(i)$ is true for all $i < n$ then we will show that $P(n)$ is also true. $P(n)$ is $T(n) \leq c_2 \cdot n \cdot \lg n$. If $P(i)$ is true for all $i < n$, since $n/2 < n$ then $P(n/2)$ is also true. Then by the inductive assumption for $i = n/2$ we have $T(n/2) \leq c_2 (n/2) \lg (n/2)$. We then use the recurrence and utilize the inductive hypothesis for $T(n/2)$

$$\begin{aligned} T(n) &= 2T(n/2) + n \\ &= 2 \overbrace{T(n/2)}^{P(n/2):\ T(n/2) \leq c_2 (n/2) \lg (n/2)} + n \\ &\leq 2c_2 (n/2) \lg (n/2) + n = c_2 n \lg n - c_2 n + n \end{aligned}$$

To complete the inductive step we must show that the rightmost expression $c_2 n \lg n - c_2 n + n$ is at most $c_2 n \lg n$. For that to be true i.e $c_2 n \lg n - c_2 n + n \leq c_2 n \lg n$ we need $-c_2 n + n < 0$ i.e. $c_2 \geq 1$.
Thus if we choose a $c_2 = 1$ which is ≥ 1 and for the n_0 of the base case i.e. $n_0 = 1$ we have that

$$T(n) \leq n \lg n \quad \forall n \geq 1.$$

Can we do better in the guessing game? Let us try a lower bound.

B. Guess Step: $T(n) = \Omega(n \lg n)$. We guess $T(n) = \Omega(n \lg n)$, i.e. $\exists$ pos. constant $c_1, n_0 : T(n) \geq c_1 n \lg n$, $\forall n \geq n_0$.
B.0 Check Step (Strong Induction). We shall prove our claim by using induction. The $P(n)$ is $T(n) \geq c_1 n \lg n$ for arbitrarily large values of $n \geq n_0$, for some positive constant n_0.
B.1. Base Case of Induction. We show $P(1)$ is true: $T(1) \geq c_1 \cdot 1 \cdot \lg 1$. Since $T(1) = 0$, it is true that $0 = T(1) \geq c_1 \cdot 1 \cdot \lg 1 = 0$ for all choices of a positive and constant c_1. Base case completed.
B.2. Inductive Step. $P(1) \wedge \ldots \wedge P(n/2) \wedge \ldots \wedge P(n-1) \Rightarrow P(n)$. We show that if $P(i)$ is true for all $i < n$ then we will show that $P(n)$ is also true. $P(n)$ is $T(n) \geq c_1 \cdot n \cdot \lg n$. If $P(i)$ is true for all $i < n$, since $n/2 < n$ then $P(n/2)$ is also true. Then by the inductive assumption for $i = n/2$ we have $T(n/2) \geq c_1 (n/2) \lg (n/2)$. We then use the recurrence and utilize the inductive hypothesis for $T(n/2)$.

$$\begin{aligned} T(n) &= 2T(n/2) + n \\ &= 2 \overbrace{T(n/2)}^{P(n/2):\ T(n/2) \geq c_1 (n/2) \lg (n/2)} + n \\ &\geq 2c_1 (n/2) \lg (n/2) + n = c_1 n \lg n - c_1 n + n \end{aligned}$$

To complete the inductive step we must show that the rightmost expression $c_1 n\lg n - c_1 n + n$ is at least $c_1 n\lg n$. For that to be true i.e $c_1 n\lg n - c_1 n + n \geq c_1 n\lg n$ we need $-c_1 n + n > 0$ i.e. $c_1 \leq 1$.
Thus if we choose a $c_1 = 1$ which is ≤ 1 and for the n_0 of the base case i.e. $n_0 = 1$ we have that

$$T(n) \geq n\lg n \quad \forall n \geq 1.$$

Well we have just proven that $T(n) \geq n\lg n$. We also proved (previous page) that $T(n) \leq n\lg n$. Both of them prove that $T(n) = n\lg n$! (The last mark is an exclamation mark, not a factorial!)
Thus while our objective was to prove a tight asymptotic bound the derivation of the three constants n_0, c_1, c_2 allowed us to determine an exact bound.
C. Guess Step : $T(n) = An\lg n + Bn + C$. Let us guess an answer that contains a log-linear, a linear and a constant term. Let's try to compute the constant values A, B and C. Only condition is that A should be positive.
C.1. Base Case utilization. We have $T(1) = 0$ substituting 1 for n in $T(n) = An\lg n + Bn + C$ we get $0 = T(1) = A \cdot 1 \cdot \lg 1 + B \cdot 1 + C$ i.e. $B + C = 0$.
C.2. Recurrence utilization. Since $T(n) = 2T(n/2) + n$ we have that

$$\begin{aligned} T(n) &= 2T(n/2) + n \\ An\lg n + Bn + C &= 2(An/2\lg(n/2 + Bn/2 + C) + n \\ An\lg n + Bn + C &= An\lg n + (B - A + 1)n + 2C \end{aligned}$$

In the last equality the first term of both sides cancels out. We then equate the coefficients of n and the constant terms and solve for A, B, C also utilizing the base case $B + C = 0$.
The constant terms C and $2C$ should be equal to each other for $C = 0$. Since $B + C = 0$ we also have $B = 0$.
The linear term coefficients B and $(B - A + 1)$ should be equal to each other $B = B - A + 1$. This results to an $A = 1$. Thus the guessed function $T(n) = An\lg n + Bn + C$ gets resolved to having $A = 1$ and $B = C = 0$ leading to the same result with the iteratiion or previous substitution method. No need to explicitly do induction. ■

Exercise 482.

Solve the following recurrence: $T(n) = T(n/3) + T(2n/3) + 7$.

Solution.
If we try to prove that $P(n) : T(n) \leq cn$ utilizing $P(n/3)$ and $P(2n/3)$ we end up with proving

$$T(n) \leq cn/3 + c2n/3 + 7 \leq cn + 7$$

not $T(n) \leq cn$. The $+7$ is causing us problems. How do we fix this **"tiny"** problem? By subtracting something smaller than the high order term from the solution i.e. try for a solution a $T(n) \leq cn - a$. ($cn - a$ MUST be positive as well i.e. $n > a/c$). We then get

$$T(n) \leq \overbrace{cn/3 - a}^{P(n/3)} + \overbrace{2cn/3 - a}^{P(2n/3)} + 7 \leq cn - a + (7 - a).$$

In order to prove that $T(n) \leq cn - a$ we must show that $T(n) \leq cn - a + (7 - a) \leq cn - a$ i.e. $7 - a \leq 0$. For the assumption to hold it suffices that $a \geq 7$.
Remark. What if we try to solve $T(n) = T(n/2) + T(n/3) + 10n$? Guess $T(n) = \Theta(n)$ (note that $1/2$ of $n/2$ plus $1/3$ of $n/3$ is $5/6 < 1$).
Remark. What if we try to solve $T(n) = T(2n/3) + T(n/3) + 10n$? Guess $T(n) = O(n \lg n)$ as $2/3 + 1/3 = 1$. $T(n) = \Theta(n \lg n)$ is the tightest possible bound.
Remark. What if we try to solve $T(n) = T(4n/3) + T(n/3) + 10n$? Guess $T(n) = O(n^\alpha)$ or $T(n) = \Theta(n^\alpha)$ for some constant $\alpha > 1$.
Remark. Useful Tricks: Change of variables might help. Remember to derive the correct boundary conditions when changing variables. $T(n) = T(\sqrt{n}) + \lg n$. Change variables $H(m) = T(2^m)$. ■

Exercise 483.

Solve the following recurrence using the substitution method: $T(n) = T(n/3) + T(n/5) + 90n$, $T(1) = 45$.

Solution.

1. Showing Ω directly, not by induction.

For $n = 1$, $T(n)$ is given by the base case and for all other natural values, $T(n)$ is given by the recurrence. We observe that because $T(n) = T(n/3) + T(n/5) + 90n$, we have obviously $T(n) = T(n/3) + T(n/5) + 90n \geq 90n$. Therefore let us try to establish $Q(n) : T(n) \geq 90n$. $Q(n)$ is true for all $n > 1$. Is this, however, true for $n = 1$ as well?

For $n = 1$ using the base case we have $45 = T(1)$. However establishing $Q(1)$ is true is not possible. $45 = T(1) \geq 90 \cdot 1$ is false! Therefore we can not show that $T(n) \geq 90n$ for all $n \geq 1$!

This is easily fixable by observing that $T(n) = T(n/3) + T(n/5) + 90n \geq 45n$. Therefore we shall prove that $P(n) : T(n) \geq 45n$ is true for all $n \geq 1$ by also showing that it is true for $n = 1$, given that by the recurrence $T(n) \geq 45n$, for $n > 1$. Using the base case for $n = 1$ $45 = T(1) \geq 45 \cdot 1$ we derive $P(n)$ true for all $n \geq 1$.

Therefore $T(n) \geq 45n$ for all $n \geq 1$, i.e. there exist positive constant $c_1 = 45 > 0$ and $n_1 = 1$ such that for all $n \geq n_1$. This shows that $T(n) = \Omega(n)$.

2. Showing $O(n)$ using induction.

We show that $T(n) = O(n)$ using strong induction, i.e we are going to show that "**there exist positive constants n_2 and c_2 such that $T(n) \leq c_2 n$ for all $n \geq n_2$** ". We shall call $P(n)$ the $T(n) \leq c_2 n$, and show $P(n)$ true for all $n \geq n_2$ for some postivie n_2, c_2. **2's Base Case.** This is for $n = 1$. We are going to show $P(1)$ is true. For $n = 1$ we have $T(1) = 45$. By $P(1)$, $45 = T(1) \leq_? c_2 \cdot 1$ is true as long as $c_2 \geq 45$. Therefore $P(1)$ is true for all $n \geq 1$, as long as $c_2 \geq 45$. Our current values for c_2, n_2 are ≥ 45 and 1 respectively.

2's (Strong) Inductive Hypothesis. We are going to use strong induction i.e. we are going to establish that "if P(j) is true for all $j = n_2, \ldots, n-1$, then $P(n)$ will also be true".

This is equivalent (since $n_2 = 1$) to "If $T(j) \leq c_2 j$ for all $j = 1, \ldots, n-1$, then $T(n) \leq c_2 n$".

2's (Strong) Inductive Step. We show the inductive step using the recurrence. By the Inductive Hypothesis since $j = n/3 < n$ and also $j = n/5 < n$, we have that $P(n/3)$ and $P(n/5)$ are both true and thus $T(n/3) \leq c_2(n/3)$ and $T(n/5) \leq c_2(n/5)$. We apply the inductive hypothesis twice in the first and second terms of the first equation below, whereas our objective is to establish the inequality on the fifth line.

$$\begin{aligned} T(n) &= T(n/3) + T(n/5) + 90n \\ &= c_2(n/3) + T(n/5) + 90n \\ &= c_2(n/3) + c_2(n/5) + 90n \\ &= c_2(8n/15) + 90n \\ &\leq_? c_2 n \end{aligned}$$

For the latter to be true we need to have $c_2 \geq 200$ as is shown in detail below.

$$\begin{aligned} c_2(8n/15) + 90n &\leq_? c_2 n \\ 90n &\leq_? 7c_2 n/15 \\ 1350 &\leq_? 7c_2 \\ 200 &\leq_? c_2 \end{aligned}$$

Therefore the inductive step is true for all n and $c_2 \geq 200$. Again, 200 is not the best possible constant, $1350/7$ is less than 200, more accurate but more cumbersome. This ($c_2 \geq 200$) supersedes the $c_2 \geq 45$ of the base case;

the induction is valid for all $n \geq 1$ and $c_2 \geq \max(45, 200) = 200$. Therefore there exist $c_2 = 200$ and $n_2 = 1$ such that $T(n) \leq c_2 n$ for all $n \geq n_2$. Thus $T(n) = O(n)$.
We proved O and Ω i.e. $T(n) = \Theta(n)$. ■

Exercise 484.

Iteration or Recursion tree method. If the method is done graphically it is sometimes called the **recursion tree method**. The traditional name is **iteration method**.
Solve exactly $T(n) = 2T(n/2) + n, \qquad T(1) = 0$. You may assume that n is a power of two.

Solution.
In order to compute $T(n)$ we need to establish $T(n/2)$. Likewise in order to compute $T(n/2)$ we need to establish $T(n/4)$. We iterate i iterations until $T(n/2^i)$ with the intent of making $n/2^i$ equal to the base case value 1. If $n/2^i = 1$ then we know that $T(n/2^i) = T(1) = 0$. For $n/2^i = 1$ we have $n = 2^i$ and solving for i we get $i = \lg n$. Thus unfolding the recurrence involves $i = \lg n$ iterations!

$$n \to n/2^1 \to n/2^2 \to n/2^3 \to \ldots n/2^i \ldots \qquad \text{until } n/2^i \text{ is base case1.}$$

$$\begin{aligned}
T(n) &= 2T(n/2) + n = 2\overbrace{\left(2T(n/2^2) + n/2\right)}^{T(n/2)} + n \\
&= 2^2 T(n/2^2) + 2n = 2^2\overbrace{\left(2T(n/2^3) + n/2^2\right)}^{T(n/2^2)} + 2n \\
&= 2^3 T(n/2^3) + 3n \\
&\ldots \\
&= 2^i T(n/2^i) + i \cdot n \quad \textbf{Now, Substitute} \quad i = \lg n \\
&= 2^{\lg n} \cdot T(n/2^{\lg n}) + \lg n \cdot n = n \cdot T(1) + \lg n \cdot n = n \lg n
\end{aligned}$$

- Keep track of number of iterations/depth of recursion tree.
- Keep track of sum of terms per iteration/level of recursion tree.
- Sometimes, the two previous steps and experience allow us to guess the solution correctly. We can then stop the solution with this method and switch to the substitution method instead.

■

Exercise 485.
Solve exactly $T(n) = 2T(n/2) + n$, $T(2) = 1$. Assume n is a power of 2.

Solution.
Same recurrence but different formulation. The recurrence becomes formula (A1). We then substitute $n/2$, $n/2^2$, $n/2^3$, etc for n in (A1). This way we generate (A2), (A3), through (Ai). (Note that $T(n) = 2T(n/2) + n$ gets rewritten as $T(n/2^0) = 2T(n/2^1) + n/2^0$.

$$\begin{array}{lll} A1 & is & T(n/2^0) = 2T(n/2^1) + n/2^0 \\ A2 & is & T(n/2^1) = 2T(n/2^2) + n/2^1 \\ A3 & is & T(n/2^2) = 2T(n/2^3) + n/2^2 \\ A4 & is & T(n/2^3) = 2T(n/2^4) + n/2^3 \\ & \dots & \\ Ai & is & T(n/2^{i-1}) = 2T(n/2^i) + n/2^{i-1} \end{array}$$

Moving forward, we use (A2) to rewrite (A1) then (A3), then (A4) all the way to (Ai).

$$\begin{array}{rcl} T(n) & = & 2T(n/2) + n =_{(A2)} 2\left(2T(n/2^2) + n/2^1\right) + n/2^0 = 2^2 T(n/2^2) + 2n \\ & = & 2^2 T(n/2^2) + 2n =_{(A3)} 2^2\left(2T(n/2^3) + n/2^2\right) + 2n = 2^3 T(n/2^3) + 3n \\ & = & 2^3 T(n/2^3) + 3n =_{(A4)} 2^3\left(2T(n/2^4) + n/2^3\right) + 3n = 2^4 T(n/2^4) + 4n \\ & \dots & \\ & = & 2^i T(n/2^i) + i \cdot n \end{array}$$

Thus $T(n) = 2^i T(n/2^i) + i \cdot n$. We want to reach the base case i.e. $T(n/2^i) = T(1)$. This implies $n/2^i = 1$ i.e. $i = \lg n$. Continuing from where we stopped using $i = \lg n$ we have the following.

$$\begin{array}{rcl} T(n) & = & 2^i \cdot T(n/2^i) + i \cdot n \\ & = & 2^{\lg n} \cdot T(n/2^{\lg n}) + \lg n \cdot n \\ & = & n \cdot T(1) + n \lg n = n \cdot 0 + n \lg n \\ & = & n \lg n \end{array}$$

We conclude that $T(n) = n \lg n$ as needed. ■

Exercise 486.

Example redone. Solve exactly $T(n) = 2T(n/2) + n$, $T(2) = 1$. Assume n is a power of 2.

Solution.
Same recurrence but different re-formulation.

$$\begin{array}{rcl} A1 & is & T(n) = 2T(n/2) + n \\ A2 & is & T(n/2^1) = 2T(n/2^2) + n/2^1 \\ A3 & is & T(n/2^2) = 2T(n/2^3) + n/2^2 \\ A4 & is & T(n/2^3) = 2T(n/2^4) + n/2^3 \\ & \cdots & \\ Ai & is & T(n/2^{i-1}) = 2T(n/2^i) + n/2^{i-1} \end{array}$$

We notice that the left-hand side appears in the right-hand side of the previous equation. Thus we can multiply each equation with some quantity and then add up the resultin equations. For the base case we reutilize that $T(n/2^i) = T(1)$ implies $i = \lg n$ as before.

$$\begin{array}{rcl} T(n) & = & 2T(n/2) + n \\ 2^1 \times T(n/2^1) & = & 2^1 \times \left(2T(n/2^2) + n/2^1\right) \\ 2^2 \times T(n/2^2) & = & 2^2 \times \left(2T(n/2^3) + n/2^2\right) \\ 2^3 \times T(n/2^3) & = & 2^3 \times \left(2T(n/2^4) + n/2^3\right) \\ & \cdots & \\ 2^{i-1} \times T(n/2^{i-1}) & = & 2^{i-1} \times \left(2T(n/2^i) + n/2^{i-1}\right) \end{array}$$

We then have

$$\begin{array}{rcl} T(n) & = & 2T(n/2) + n \\ 2^1 \cdot T(n/2^1) & = & 2^2 T(n/2^2) + n \\ 2^2 \cdot T(n/2^2) & = & 2^3 T(n/2^3) + n \\ 2^3 \cdot T(n/2^3) & = & 2^4 T(n/2^4) + n \\ & \cdots & \\ 2^{i-1} \cdot T(n/2^{i-1}) & = & 2^i T(n/2^i) + n \end{array}$$

All but the first term of the left hand side cancel out. The n terms of the right hand side are retained. The inner terms of the right hand side cancel out except for the last term. The end result is that $T(n) = 2^i T(n/2^i) + n$. We conclude that $T(n) = n \lg n$ as needed. ■

Exercise 487.
Solve exactly $T(n) = 2T(n/2) + n$ where $T(4) = 12$. You may assume n is a power of 2.

Solution.

$$\begin{array}{rcl} A1 & is & T(n/2^0) = 2T(n/2^1) + n/2^0 \\ A2 & is & T(n/2^1) = 2T(n/2^2) + n/2^1 \\ A3 & is & T(n/2^2) = 2T(n/2^3) + n/2^2 \\ A4 & is & T(n/2^3) = 2T(n/2^4) + n/2^3 \\ & \cdots & \\ Ai & is & T(n/2^{i-1}) = 2T(n/2^i) + n/2^{i-1} \end{array}$$

Moving forward, we use (A2) to rewrite (A1) then (A3), then (A4) all the way to (Ai).

$$\begin{array}{rcl} T(n) & = & 2T(n/2) + n =_{(A2)} 2\left(2T(n/2^2) + n/2^1\right) + n/2^0 = 2^2T(n/2^2) + 2n \\ & = & 2^2T(n/2^2) + 2n =_{(A3)} 2^2\left(2T(n/2^3) + n/2^2\right) + 2n = 2^3T(n/2^3) + 3n \\ & = & 2^3T(n/2^3) + 3n =_{(A4)} 2^3\left(2T(n/2^4) + n/2^3\right) + 3n = 2^4T(n/2^4) + 4n \\ & \cdots & \\ & = & 2^iT(n/2^i) + i \cdot n \end{array}$$

Thus $T(n) = 2^iT(n/2^i) + i \cdot n$. We want to reach the base case i.e. $T(n/2^i) = T(4)$. This implies $n/2^i = 4$ i.e. $i = \lg n - 2$. Continuing from where we stopped using $i = \lg n - 2$ we have the following.

$$\begin{array}{rcl} T(n) & = & 2^i \cdot T(n/2^i) + i \cdot n \\ & = & 2^{\lg n-2} \cdot T(n/2^{\lg n-2}) + (\lg n - 2) \cdot n \\ & = & (n/4) \cdot T(4) + n(\lg n - 2) \\ & = & n/4 \cdot 12 + n(\lg n - 2) \\ & = & n \lg n + n \end{array}$$

We conclude that $T(n) = n \lg n + n$ as needed. This is the closed form solution (answer). ■

Exercise 488.
Solve exactly $T(n) = 2T(n/2) + n - 1, \qquad T(1) = 0$. You may assume n is a power of two.

Solution.
Again we shall terminate recursion for $T(n/2^i)$ when $n/2^i = 1$ and solving for i we shall get $i = \lg n$. We shall also need the geometric sequence result $2^0 + 2^1 + \ldots + 2^{i-1} = 2^i - 1$.

$$\begin{aligned}
T(n) &= 2T(n/2) + n - 1 \\
&= 2\left(2T(n/2^2) + n/2 - 1\right) + n - 1 = 2^2 T(n/2^2) + 2n - 1 - 2 \\
&= 2^2\left(2T(n/2^3) + n/2^2 - 1\right) + 2n - 1 - 2 = 2^3 T(n/2^3) + 3n - 2^0 - 2^1 - 2^2 \\
&\ldots \\
&= 2^i T(n/2^i) + i \cdot n - 2^0 - 2^1 - \ldots - 2^{i-1} = 2^i T(n/2^i) + i \cdot n - (2^i - 1) \\
&= 2^i T(n/2^i) + i \cdot n - (2^i - 1) \quad \textbf{Now, Substitute} \quad i = \lg n \\
&= 2^{\lg n} \cdot T(n/2^{\lg n}) + \lg n \cdot n - (2^{\lg n} - 1) \\
&= n \cdot T(1) + \lg n \cdot n - n + 1 = n \lg n - (n - 1)
\end{aligned}$$

The previous recurrences were generating an additive term which was the same in each iteration: n. This one generates a term that is iteration dependent: $n - 2^{i-1}$ for iteration i. Thus for $i = 1$ we get $n - 1$, for $i = 2$ we get $n - 2$, and so on.
The answer for $T(n)$ with the $(-(n-1)$ low-order term) is the one we got through the recursion tree method as well. ■

Exercise 489.

Solve exactly the recurrence $T(n) = 2T(n/2) + n$, $T(2) = 5$. Assume n is a power of 2.

Solution.
Substituting $n/2$ for n in the recurrence we get. $T(n/2) = 2T((n/2)/2) + n/2 = 2T(n/2^2) + n/2$.
Substituting $n/4 = n/2^2$ for n we get. $T(n/2^2) = 2T((n/4)/2) + n/4 = 2T(n/2^3) + n/2^2$
Substituting $n/8 = n/2^3$ for n we get. $T(n/2^3) = 2T(n/2^4) + n/2^3$
Similarly we can substitute $n/2^4, \ldots, n/2^{i-1}, \ldots$ for n to resolve the 4, ..., $i-1$-st iteration all the way to the base case, and get similarly stated recurrences. We utilize all the derived recurrences to expand $T(n)$ by observing that $T(n)$ is $2T(n/2) + n$. Then we use the derivation for $T(n/2)$ to formulate the expression in terms of $T(n/2^2)$, then $T(n/2^2)$ is expressed in terms of $T(n/2^3)$ and so on ... in terms of $T(n/2^i)$. This goes on until we reach the based case of $n = 2$ since $T(2) = 5$ is given. We establish what value of i provides the base case by equating $T(2) = T(n/2^i)$ and solving for i. The unfolding of the recurrence stops then and all the terms unfolded get combined (summed). A closed form solution for $T(n)$ can then be found.

$$\begin{aligned} T(n) &= 2T(n/2) + n \;=\; 2(2T(n/2^2) + n/2) + n \;=\; 2^2 T(n/2^2) + 2n \\ &= 2^2(2T(n/2^3) + n/2^2) + 2n \;=\; 2^3 T(n/2^3) + n + 2n \;=\; 2^3 T(n/2^3) + 3n \\ &= \ldots = 2^i T(n/2^i) + i \cdot n \end{aligned}$$

From the boundary condition $T(2) = 5$, we decide when to stop the unfolding of $T(n)$. The expansion ends at $n/2^i = 2$ since then $T(n/2^i) = T(2) = 5$. We solve for i by taking logarithms base two of both sides. Then we get that $\lg n - i = 1$ ie $i = \lg n - 1$. Then, for $i = \lg n - 1$, we get that.

$$\begin{aligned} T(n) &= 2T(n/2) + n = \ldots = 2^i T(n/2^i) + i \cdot n \\ &= 2^{\lg n - 1} T(n/2^{\lg n - 1}) + (\lg n - 1) \cdot n \\ &= (n/2)T(2) + (\lg n - 1) \cdot n = (n/2)5 + (\lg n - 1) \cdot n \\ &= (3n/2) + n \lg n \end{aligned}$$

Thus we have obtained the following solution to the recurrence: $T(n) = 3n/2 + n \lg n$. ■

Exercise 490.
Solve exactly $T(n) = aT(n/b)$ where $T(1) = 1$. a, b as in the master method.

Solution.
First we generate (A1) through (Ai) as needed.

$$\begin{aligned} T(n) &= aT(n/b) \\ \\ T(n/b^0) &= aT(n/b^1) \\ T(n/b^1) &= aT(n/b^2) \\ T(n/b^2) &= aT(n/b^3) \\ T(n/b^3) &= aT(n/b^4) \\ &\cdots \\ T(n/b^{i-1}) &= aT(n/b^i) \end{aligned}$$

As there are no additive terms above, we can multiply all the equations but the first together (The second equation is the first equation rewritten; it makes no sense to use both of them). The result is $T(n) = a^i T(n/b^i)$. We then equate $T(n/b^i)$ with the base case $T(1)$ and solving for i we determine the number of iteration needed. $n/b^i = 1$ implies $i = \lg n / \lg b$.

$$\begin{aligned} T(n) &= a^i T(n/b^i) \\ &= a^{\lg n/\lg b} T(n/b^{\lg n/\lg b}) \\ &= 2^{\lg a \lg n/\lg b} T(n/2^{\lg b \lg n/\lg b}) \\ &= n^{\lg a/\lg b} T(n/2^{\lg n}) \\ &= n^{\lg_b a} T(n/n) \\ &= n^{\lg_b a} T(1) \\ &= n^{\lg_b a}. \end{aligned}$$

Thus $T(n) = n^{\lg_b a}$. ■

Exercise 491.
Solve the recurrence $T(n) = 8T(n/2) + n$ using the iteration/recursion tree method. Assume that $T(1) = 5$.

Solution.

$$\begin{aligned}
T(n) &= 8T(n/2)+n \\
&= 8(8T(n/2^2)+n/2)+n \\
&= 8^2T(n/2^2)+8n/2+n \\
&= 8^2T(n/2^2)+(8/2)^1n+(8/2)^0n \\
&= 8^2(8T(n/2^3)+n/2^2)+(8/2)^1n+(8/2)^0n \\
&= 8^3T(n/2^3)+8^2n/2^2+(8/2)^1n+(8/2)^0n \\
&= 8^3T(n/2^3)+(8/2)^2n+(8/2)^1n+(8/2)^0n \\
&= \dots \\
&= 8^iT(n/2^i)+(8/2)^{i-1}n+\dots+(8/2)^1n+(8/2)^0n
\end{aligned}$$

Again the boundary case is $T(1) = 5$. We set $n/2^i = 1$, ie $i = \lg n$. Then for $i = \lg n$, $T(n/2^i) = T(1) = 5$. We therefore get for $i = \lg n$.

$$\begin{aligned}
T(n) &= 8T(n/2)+n \\
&= 8^iT(n/2^i)+(8/2)^{i-1}n+\dots+(8/2)^1n+(8/2)^0n \\
&= 8^{\lg n}T(n/2^{\lg n})+\frac{(8/2)^{\lg n}-1}{8/2-1}\cdot n \\
&= 2^{3\lg n}T(1)+\frac{(4)^{\lg n}-1}{3}\cdot n \\
&= n^3T(1)+\frac{(4)^{\lg n}-1}{3}\cdot n \\
&= 5n^3+\frac{n^2-1}{3}n
\end{aligned}$$

To verify our calculations we observe that $T(1) = 5+(1-1)/3 = 5$ and

$$\begin{aligned}
T(n) &= 8T(n/2)+n \\
&= 8(5(n/2)^3+\frac{(n/2)^2-1}{3}n)+n \\
&= 5n^3+\frac{n^2-1}{3}n \\
&= T(n),
\end{aligned}$$

ie the recurrence is verified. ■

Exercise 492.
Use the iteration method to solve the recurrence, $T(n) = 4T(n/2) + n$, $T(4) = 6$.

Solution.

$$\begin{aligned}
T(n) &= 4T(n/2) + n \\
&= 4\left(4T\left(\frac{n}{2^2}\right) + \left(\frac{n}{2}\right)\right) + n \\
&= 4^2 T\left(\frac{n}{2^2}\right) + \frac{4^1 n}{2^1} + n \\
&= 4^2\left(4T\left(\frac{n}{2^3}\right) + \left(\frac{n}{2^2}\right)\right) + 2^1 n + 2^0 n \\
&= 4^3 T\left(\frac{n}{2^3}\right) + 2^2 n + 2^1 n + 2^0 n \\
&= \ldots \\
&= 4^i T\left(\frac{n}{2^i}\right) + 2^{i-1} n + \ldots + 2^1 n + 2^0 n \\
&= 4^i T\left(\frac{n}{2^i}\right) + n\left(2^i - 1\right)
\end{aligned}$$

The base case is $T(4) = 6$. We set $n/2^i = 4$, i.e. $n = 2^{i+2}$. If we solve this for i, we get $i = \lg n - 2$. Then $T(n/2^i) = T(4) = 6$. In addition, $2^i = n/4$, $4^i = 2^i \cdot 2^i = n/4 \cdot n/4 = n^2/16$. Therfore the formula for $T(n)$ becomes.

$$\begin{aligned}
T(n) &= 4^i T\left(\frac{n}{2^i}\right) + n\left(2^i - 1\right) \\
&= (n^2/16)T(4) + n(n/4 - 1) \\
&= 3n^2/8 + n^2/4 - n \\
&= 5n^2/8 - n
\end{aligned}$$

■

Exercise 493.

Use mathematical induction to show that when n is an exact power of two, the solution of the recurrence

$$T(n) = \begin{cases} 8 & \text{if } n = 2 \\ 2T(n/2) + n & \text{if } n = 2^k, \text{ for } k > 1 \end{cases}$$

is $T(n) = n\lg n + 3n$.

Solution.
We show the claim by induction. We first form proposition $P(n)$ that is to be proved.
Proposition $P(n)$: $T(n) = 3n + n\lg n$.
Since we only need to show $P(n)$ not for all n but only for n being a power of two, we can transform $P(n)$ into a $Q(k)$ such that $Q(k) = P(2^k)$ using $n = 2^k$ and $k \geq 1$. $Q(k)$ is thus given by the following expression.

$$Q(k) \quad : \quad T(2^k) = 3 \cdot 2^k + k \cdot 2^k = (k+3) \cdot 2^k.$$

The recurrence $T(n) = 2T(n/2) + n$ can also be written in terms of k since $n = 2^k$ as

$$T(2^k) = 2T(2^{k-1}) + 2^k.$$

We show $Q(k)$ by induction on k.
Base case ($Q(1)$). For $k = 1$ and subsituting for $k = 1$ in $Q(k)$ we get that $T(2^1) = (1+3)2^1 = 8$. This is true by the way of the base case $T(2) = 8$ that is given.
Inductive step $Q(k-1) \Longrightarrow Q(k)$. $Q(k-1)$ is equivalent to $T(2^{k-1})$ i.e. $T(2^{k-1}) = (k+2)2^{k-1}$. We expect to establish the implication involving $Q(k)$ i.e. $T(2^k) = (k+3) \cdot 2^k$, from $Q(k-1)$.
We use the recurrence $T(2^k) = 2T(2^{k-1}) + 2^k$ and in it we substitute for the $T(2^{k-1})$ as obtained from $Q(k-1)$.

$$\begin{aligned} T(n) = T(2^k) &= 2 \cdot T(2^{k-1}) + 2^k \\ &\overset{Q(k-1)}{=} 2 \cdot ((k+2) \cdot 2^{k-1}) + 2^k \\ &= k \cdot 2^k + 2 \cdot 2^k + 2^k \\ &= k \cdot 2^k + 3 \cdot 2^k = (k+3) \cdot 2^k \end{aligned}$$

We have thus proved that $T(2^k) = (k+3) \cdot 2^k$ i.e. $Q(k)$ is true. This completes the induction and shows also $P(n)$. ■

Exercise 494.

Use mathematical induction to show that when n is an exact power of two, the solution of the recurrence

$$T(n) = \begin{cases} 4 & \text{if } n = 1 \\ 2T(n/2) + 4n & \text{if } n = 2^k, \text{ for } k \geq 1 \end{cases}$$

is $T(n) = 4n \lg n + 4n$.

Solution.
We show the claim by induction. $P(n)$ is $T(n) = 4n + 4n \lg n$. Since we only need to show $P(n)$ not for all n but only for n being a power of two, we can transform $P(n)$ into $Q(k) = P(2^k)$ using $n = 2^k$. $Q(k)$ is thus $T(2^k) = 2^{k+2} + k2^{k+2} = (k+1)2^{k+2}$. The recurrence $T(n) = 2T(n/2) + 2n$ can also be written in terms of k as $T(2^k) = 2T(2^{k-1}) + 2^{k+2}$. We show $Q(k)$ by induction on k.
Base case. For $k = 0$, we have $n = 2^0 = 1$. We substitute 0 for k in $T(2^k)$. Thus $T(2^k) = T(1)$ which is 4 by the base case of the recurrence that was given to us. Starting from the right side of $T(2^k)$ which is $(k+1)2^{k+2}$ for $k = 0$ we get $(0+1)2^{0+2} = 4$ as well. Therefore we have just shown that $Q(0)$ is true because both sides of $T(2^k) = (k+1)2^{k+2}$ give 4 for $k = 0$.
Inductive Hypothesis. We show that if $Q(k-1)$ is true, then $Q(k)$ will also be true.
Inductive Step. If $Q(k-1)$ is true, then $T(2^{k-1}) = k2^{k+1}$. We expect to establish that $Q(k)$ is true, i.e. $T(2^k) = (k+1) \cdot 2^{k+2}$. We use the recurrence $T(2^k) = 2T(2^{k-1}) + 2^{k+2}$ to show that $Q(k)$ is true then. We start from the recurrence above and we substitute there for $T(2^{k-1})$ the expression obtained by the inductive hypothesis for $Q(k-1)$.

$$\begin{aligned} T(n) = T(2^k) &= 2T(2^{k-1}) + 2^{k+2} \\ &= 2(k2^{k+1}) + 2^{k+2} \\ &= k2^{k+2} + 2^{k+2} \\ &= (k+1) \cdot 2^{k+2} \end{aligned}$$

We have thus proved that $T(2^k) = (k+1)2^{k+2}$. We express the answer as a function of n now. Since $n = 2^k$ we have that

$$\begin{aligned} T(2^k) &= (k+1) \cdot 2^{k+2} \\ T(n) &= (\lg n + 1) \cdot n \cdot 4 \\ T(n) &= 4n \lg n + 4n \end{aligned}$$

We have thus established that $T(n) = 4n \lg n + 4n$ for all $n = 2^k$, k non-negative integer. ■

Exercise 495.

Use mathematical induction to show that when n is an exact power of two, the solution of the recurrence

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ 2T(n/2) + n & \text{if } n = 2^k, \text{ for } k > 1 \end{cases}$$

is $T(n) = n \lg n + n$.

Solution.
We show the claim by induction. The $P(n)$ of induction is $T(n) = n + n \lg n$. Since we only need to show $P(n)$ not for all n but only for n being a power of two, we can transform $P(n)$ into $Q(k) = P(2^k)$ using $n = 2^k$ and then the expression for $T(n)$ to be proved becomes $T(2^k) = 2^k + k2^k = (k+1)2^k$.
Base case. For $k = 0$, we have $n = 2^0 = n = 1$ we have that $T(1) = 1 \lg 1 + 1 = 1$. We also have by the boundary case that $T(1) = 1$. Therefore the expression for $T(n)$ is satisfied by the solution $T(n) = n + n \lg n$ for $k = 0$ (equivalently, $n = 1$).
Inductive step. Let $n = 2^k$. We assume that if $Q(k-1)$ is true then we will show that $Q(k)$ is also true. If $Q(k-1)$ is true $T(2^{k-1}) = 2^{k-1} + (k-1)2^{k-1}$.
We also have by the recurrence that $T(n) = 2T(n/2) + n$ i.e. $T(2^k) = 2T(2^{k-1}) + 2^k$. We start from the recurrence above and we substitute there for $T(2^{k-1})$ the expression obtained by the inductive hypothesis for $Q(k-1)$.

$$\begin{aligned} T(n) = T(2^k) &= 2T(2^{k-1}) + 2^k \\ &= 2(2^{k-1} + (k-1)2^{k-1}) + 2^k \\ &= 2^k + (k-1)2^k + 2^k \\ &= 2^k + k2^k \\ &= n + n \lg n \end{aligned}$$

This completes the induction as we have shown $P(k)$ (or equivalently $P(n)$) assuming that $P(k-1)$ was true (or equivalently $P(n/2)$). ■

Exercise 496.
Show that for any $n \geq 0$

$$\sum_{i=0}^{i=n} i^2 = n(n+1)(2n+1)/6$$

Solution.
Use induction. ■

Exercise 497.
Show by induction that

$$\sum_{i=0}^{n} 36i^2 = 6n(n+1)(2n+1)$$

Solution.
Let us call $P(n)$ the proposition $\sum_{i=0}^{n} 36i^2 = 6n(n+1)(2n+1)$. The inductive proof follows.
Base case. We show that $P(0)$ is true. Indeed for $P(0)$ its left hand side works as $\sum_{i=0}^{0} 16i^2 = 0$ and its right hand side works as $6 \cdot 0 \cdot (0+1) \cdot (2 \cdot 0+1) = 0$. Both sides are equal thus $P(0)$ is true.
Inductive Hypothesis. We show that if $P(n)$ is true then $P(n+1)$ is true. More explicitly, we show that if $\sum_{i=0}^{n} 36i^2 = 6(n)(n+1)(2n+1)$ is true, then $\sum_{i=0}^{n+1} 36i^2 = 6(n+1)(n+2)(2n+3)$ is also true.
Inductive Step. We start with stating the left hand side of $P(n+1)$ decomposing it into the left hand side of $P(n)$ plus other terms and then substituting for the right hand side of $P(n)$ to derive eventually the right hand size of $P(n+1)$.

$$\begin{aligned} P(n) &: \sum_{i=0}^{n} 36i^2 = 6n(n+1)(2n+1) \\ P(n+1) &: \sum_{i=0}^{n+1} 36i^2 = 6(n+1)(n+2)(2n+3) \end{aligned}$$

$$\begin{aligned} \sum_{i=0}^{n+1} 36i^2 &= \sum_{i=0}^{n} 36i^2 + 36(n+1)^2 \\ &= 6n(n+1)(2n+1) + 36(n+1)(n+1) \\ &= 6(n+1)(n(2n+1) + 6(n+1)) \\ &= 6(n+1)(2n^2 + n + 6n + 6) \\ &= 6(n+1)(2n^2 + 7n + 6) \\ &= 6(n+1)(n+2)(2n+3) \end{aligned}$$

We have just proven that if $P(n)$ is true, then $P(n+1)$ will be true. This completes the induction. Therefore $P(n)$ is true for all $n \geq 0$. ■

Exercise 498.
Show by induction that

$$\sum_{i=0}^{n} i^3 = n^2(n+1)^2/4$$

Solution.
Let us call $P(n)$ the proposition $\sum_{i=0}^{n} i^3 = n^2(n+1)^2/4$.
Base case. We show that $P(0)$ is true. Indeed for $P(0)$ its left hand side works as $\sum_{i=0}^{0} i^3 = 0$ and its right hand side works as $0^2(0+1)^2/4 = 0$. Both sides are equal and equal to zero thus $P(0)$ is true.
Inductive Step. Suppose than $P(n-1)$ is true. We will then show than $P(n)$ is true, i.e. we are going to prove that if $\sum_{i=0}^{n-1} i^3 = (n-1)^2(n)^2/4$ is true, then $\sum_{i=0}^{n} i^3 = n^2(n+1)^2/4$ is also true. We start by stating the inductive hypothesis $P(n-1)$ and then by deriving $P(n)$ by manipulating the terms of the sum as follows.

$$\begin{aligned}
\sum_{i=0}^{n-1} i^3 &= (n-1)^2(n)^2/4 \\
\sum_{i=0}^{n-1} i^3 + n^3 &= (n-1)^2(n)^2/4 + n^3 \\
\sum_{i=0}^{n} i^3 &= (n-1)^2(n)^2/4 + n^3 \\
\sum_{i=0}^{n} i^3 &= (n^4 - 2n^3 + n^2)/4 + 4n^3/4 \\
\sum_{i=0}^{n} i^3 &= (n^4 + 2n^3 + n^2)/4 \\
\sum_{i=0}^{n} i^3 &= n^2(n+1)^2/4
\end{aligned}$$

Therefore starting from $P(n-1)$ and adding to both sides of the equation n^3 we have shown the validity of $P(n-1)$. This completes the proof of the inductive step and the induction.
Therefore $P(n)$ is true for all $n \geq 0$. ■

Exercise 499.
For the recurrence relation:

$$T(n) = 32 \cdot T(\frac{n}{4}) + 16n^2$$
$$T(4) = 128$$

compute $T(n)$ exactly for n a power of 4.

Solution.
Expanding out the recurrence, we get

$$\begin{aligned}
T(n) &= 16n^2 + 32(16(\frac{n}{4})^2 + 32(16(\frac{n}{16})^2 + \ldots + 32(16(\frac{n}{\frac{n}{16}})^2 + 32T(\frac{n}{\frac{n}{4}}))\ldots)) \\
&= 16n^2 + (32)(16)(\frac{n}{4})^2 + \ldots + (32^i)(16)(\frac{n}{4^i})^2 + \ldots + (32^{(\log_4 n)-2})(16)(\frac{n}{4^{(\log_4 n)-2}})^2 + 32^{(\log_4 n)-1}T(4) \\
&= \sum_{i=0}^{(\log_4 n)-2} \frac{32^i 16}{4^{2i}} n^2 + 32^{(\log_4 n)-1}T(4) \\
&= 16n^2 \sum_{i=0}^{(\log_4 n)-2} 2^i + 32^{\lg_4 n}4 \\
&= 16n^2(2^{(\log_4 n)-1} - 1) + n^{\lg_4 32}4 \\
&= 8n^2 2^{\log_4 n} - 16n^2 + 4n^{2.5} \\
&= 12n^{2.5} - 16n^2
\end{aligned}$$

■

Exercise 500.

Solve the following recurrences; make your bounds as tight as possible. Use strong induction if necessary.

$$T(n) \quad = \quad T(n/2)+T(n/3)+32n, \quad \text{where} \quad T(1)=100.$$

Solution. Since $T(n)=T(n/2)+T(n/3)+32n$ we have for $n=1$ from the base case that $T(1)=100$. We show that $T(n)=\Omega(n)$. By the recurrence we have that $T(n)\geq 32n$ for all $n>1$ since $T(n/2),T(n/3)$ are non-negative and the other term is $32n$. In addition, $T(1)=100\geq 32\cdot 1=32$. Therefore $T(n)\geq 32n$ for $n=1$ as well. Therefore $T(n)=\Omega(n)$.
We now prove that $T(n)=O(n)$ or equivalently that there exist positive constants n_2,c_2 such that $T(n)\leq c_2 n$ for all $n\geq n_2$. Proof is by induction on n.
Base Case. $T(1)=100\leq c_2\cdot 1$ is true provided that $c_2\geq 100$. This completes the base case.
Inductive Step. Let $T(i)\leq c_2 i$ be true for all $1\leq i<n$. We want to show that $T(n)\leq c_2 n$. Since $T(i)\leq c_2 i$ for all $i<n$, it is also true for $i=n/2$ and $i=n/3$ as well, and thus by the inductive hypothesis we obtain $T(n/2)\leq c_2 n/2$ and $T(n/3)\leq c_2 n/3$. Then we use these two inequalities in the recurrence to obtain.

$$\begin{aligned} T(n) &= T(n/2)+T(n/3)+32n \\ &\leq c_2 n/2+c_2 n/3+32n \\ &= 5c_2 n/6+32n \end{aligned}$$

In order to complete the inductive step we need to prove that $T(n)\leq c_2 n$ or equivalently $5c_2 n/6+32n\leq c_2 n$. The latter is equivalent to $32n\leq c_2 n/6$ or $c_2\geq 192$. Therefore the inductive step is completed and $T(n)\leq c_2 n$ for any $c_2\geq 192$. In conclusion, combining the inductive step and the base case we obtain that $T(n)\leq c_2 n$ for all $n\geq 1$ as long as $c_2\geq 192$. We can fix $c_2=192$ and thus show that $T(n)=O(n)$.
Therefore since $T(n)=O(n)$ and $T(n)=\Omega(n)$ we obtain that $T(n)=\Theta(n)$. The constants in the definition are $1,192$ and for n_0 we can still use 1.

■

Exercise 501.

Solve the recurrence $T(n) = T(n/2) + T(2n/3) + n$, where $T(1) = \Theta(1)$.

Solution. **Attempt 1.** We try $T(n) = O(n^2)$, i.e. $T(n) \leq cn^2$ for for some constant $c, n_0 > 0$ and for all $n \geq n_0$. Using by the inductive hypothesis $T(n/2) \leq c(n/2)^2$ and $T(2n/3) \leq c(2n/3)^2$ and

$$T(n) \leq cn^2/4 + 4cn^2/9 + n = c(25/36)n^2 + n \leq cn^2$$

for large enough n and $c > 0$. The bound, however, is not the best possible one.

Attempt 2. We try $T(n) \leq cn^\alpha$ where α is a constant such that $1 < \alpha < 2$. Similarly as before,

$$T(n) \leq cn^\alpha((1/2)^\alpha + (2/3)^\alpha) + n \leq cn^\alpha$$

The latter inequality holds asymptotically for large n provided that $(1/2)^\alpha + (2/3)^\alpha < 1$. If this value of $(1/2)^\alpha + (2/3)^\alpha$ is 0.99 then $0.99cn^\alpha + n$ is indeed smaller than cn^α asymptotically for large n.

Attempt 3. We try $T(n) \leq cn^\alpha - 6n$ where α are constant such that $1 < \alpha < 2$ and $\beta > 1$. (There is no magic in choosing a 6: it relates to $1/2$ and $2/3$ after originally choosing a β instead.) Similarly as before,

$$T(n) \leq cn^\alpha((1/2)^\alpha - 6n/2 + (2/3)^\alpha) - 2n/3 + n \leq cn^\alpha - 6n$$

The latter inequality holds asymptotically for large n provided that $(1/2)^\alpha + (2/3)^\alpha < 1$. If this value of $(1/2)^\alpha + (2/3)^\alpha$ is 0.99 then $0.99cn^\alpha - 6n$ is indeed smaller than $cn^\alpha - 6n$ asymptotically for large n. A numerical computation of an α yields one in the range $[1.293, 1.294]$. The exact value is of no interest; only the fact that it is between 1 and 2!

■

Exercise 502.

There are more examples with the substitution method in the pages to follow.
Solve the following recurrences. You may assume $T(1) = \Theta(1)$, where necessary. Make your bounds as tight as possible. Use asymptotic notation to express your answers. Justify your answers.
$T(n) = T(n/4) + T(3n/4) + n\lg n$.

Solution. We use the substitution method to solve this recurrence. We guess as a solution $T(n) = O(n\lg^2 n)$ ie $T(n) \leq cn\lg^2 n$, for some positive constant $c > 0$ and for all $n \geq n_0$. We use induction to prove that this is indeed a solution to the recurrence.

$$\begin{aligned} T(n) &= T(n/4) + T(3n/4) + n\lg n \\ &\leq (cn/4)\lg^2(n/4) + (3cn/4)\lg^2(3n/4) + n\lg n \\ &= c(n/4)(\lg n - 2)^2 + c(3n/4)(\lg n + \lg 3 - 2)^2 + n\lg n \\ &= c(n/4)(\lg n - 2)^2 + c(3n/4)(\lg n)^2 + n\lg n \\ &\leq cn\lg^2 ncn + n - cn\lg n \\ &\leq cn\lg^2 n \end{aligned}$$

for any $n \geq 4$ and $c \geq 1$. We choose for boundary condition $T(2) = 2$ so that $T(n) \leq cn\lg^2 n$ would result in $2 \leq c2\lg^2 2 = 2c$, ie $c \geq 1$. Therefore for any $n \geq 4$ and any $c \geq 1$, we have $T(n) \leq cn\lg^2 n$.

■

Exercise 503.

Solve the following recurrences associated with three problems introduced in class. First name the problem associated with a recurrence and then solve the recurrence as best as you can (i.e. tightest possible bound) asymptotic or not! You may assume n is a power of 2. Formulate appropriate base case as well.

(a) $T(n) = T(n/2) + 1$.
(b) $T(n) = 2T(n/2) + (n-1)$.
(c) $T(n) = T(n-1) + (n-1)$.

Solution. (a) $T(1) = 1$, Binary search and $T(n) = \lg n + 1$ for (upper bound on) number of comparison in worst-case thus determining running time also but using a rougher analysis ($n/2$ should have been $(n-1)/2$).
(b) $T(2) = 1$, Merge sort and $T(n) = n\lg n - (n-1)$ for number of comparisons in worst-case thus determining running time also.
(c) $T(1) = 0$, Insertion sort (recursive formulation) $T(n) = n(n-1)/2$ for number of comparisons in the worst-case determining running time also.

■

Exercise 504.

Solve the recurrennce $T(n) = T(an) + T(bn) + n$ where $a + b < 1$ and $a > 0$ and $b > 0$ are constant. In the following problem, we address the case $a + b = 1$.

Solution. We could use the substitution method to solve this recurrence.

(1.) Ω part: $T(n) = \Omega(n)$ is straightforward as $T(n) = T(an) + T(bn) + n \geq n$.

(2.) O part: We show $T(n) = O(n)$ i.e. we show that there exist c, n_0 positive and constant such that $T(n) \leq cn$ for $\forall n \geq n_0$. Proof by induction. Since $an < n$ and $bn < n$ the induction hypothesis for $T(an)$ and $T(bn)$ gives

$$\begin{aligned} T(n) &= T(an) + T(bn) + n \\ &\leq c(an) + c(bn) + n \\ &= cn(a+b) + n \end{aligned}$$

For the last expression to be at most cn we need to have

$$\begin{aligned} cn(a+b) + n &\leq cn \\ n &\leq c(1-a-b)n \\ 1 &\leq c(1-a-b) \end{aligned}$$

Since $a + b < 1$ we have $1 - a - b > 0$. Therefore if we choose c such that $c > 1/(1-a-b)$ the last inequality is true and the inductive step is proven.

■

Exercise 505.

Solve the following recurrence. Use asymptotic notation to express your solution. Give a solution as asymptotically tight as possible.

$$T(n) = T(n/2) + T(3n/4) + n$$

Solution.

1st Approximation By using the substitution method we guess as a first solution $T(n) = O(n^2)$. It is easy to see that it satisfies the recurrence. However it is not the tightest possible solution.

2nd Approximation

Subsequently, we guess $T(n) = O(n^a)$ for some constant $1 < a < 2$. Then by induction

$$\begin{aligned} T(n) &= T(n/2) + T(3n/4) + n \\ &\leq c(n/2)^a + c(3n/4)^a + n \end{aligned}$$

The last expression $c(n/2)^a + c(3n/4)^a + n$ is at most cn^a provided that

$$c(n/2)^a + c(3n/4)^a + n \leq cn^a$$

which is equivalent to

$$(1/2)^a + (3/4)^a + (1/c)n^{1-a} \leq 1$$

Note that the third term of the left hand side is the only non-constant term and since $0 < a < 1$ and a is constant, it grows asymptotically to zero. We then observe by numerical methods that $(1/2)^a + (3/4)^a - 1$ for $a = 1.507$ is positive and $a = 1.508$ is negative ie the solution to the $(1/2)^a + (3/4)^a = 1$ is an a such that $1.507 < a < 1.508$. We call this solution b. For $a = 1.508$, we have that $(1/2)^a + (3/4)^a - 1$ is -0.003, therefore the expression

$$(1/2)^a + (3/4)^a + (1/c)n^{1-a} \leq 1$$

becomes

$$1 - 0.003 + (1/c)n^{1-a} \leq 1$$

and

$$(1/c)n^{1-a} \leq 0.003$$

for $c = 1$ and any $n > n_0 = 10^9$. This solution, however **is not tight**. The reason is the existence of the additive term n, which, however, is asymptotically smaller than n^a as $a > 1$.

3rd Approximation

We thus guess as a solution a variant of the 2nd approximation ie $T(n) = cn^a - bn$ for some constants $1 < a < 2$, $c, b > 0$. Then by induction

$$\begin{aligned} T(n) &= T(n/2) + T(3n/4) + n \\ &\leq c(n/2)^a - bn/2 + c(3n/4)^a - b3n/4 + n \end{aligned}$$

We examine the last expression $c(n/2)^a + c(3n/4)^a + n - b5n/4$ which is at most $cn^a - bn$ provided that (a) we first eliminate the linear terms, ie set $n - 5bn/4 = -bn$ which implies that we need to set $b = 5$. (b) Then, it simplifies into

$$c(n/2)^a + c(3n/4)^a \leq cn^a$$

which is equivalent to

$$(1/2)^a + (3/4)^a \leq 1.$$

Thus the solution for a of $(1/2)^a + (3/4)^a = 1$ which lies somewhere between 1.507 and $a = 1.508$ gives the value of the exponent a in $T(n) = \Theta(n^a)$.

Thus, a tight bound for the recurrence is $T(n) = cn^a - 5n$, where c can be any constant $c > 0$ and a is the solution of $(1/2)^a + (3/4)^a = 1$ which is a real number between 1 and 2 (or more accurately between 1.507 and 1.508 if one relies to numerical methods). **Note.** The best possible bound is the one derived from the third approximation. The other two steps are not necessary; they just show how one can successively refine a guess in the substitution method. Note also, that the numerical computations we used to find that a is between 1.507 and 1.508 are not necessary, as the exponent a can just be expressed as a solution to an equation (ie $(1/2)^a + (3/4)^a = 1$). ■

Exercise 506.

Solve the following recurrences. You may assume $T(1) = \Theta(1)$, where necessary. Make your bounds as tight as possible. Use asymptotic notation to express your answers. Justify your answers.

$$T(n) = T(n/4) + T(3n/4) + n\lg n.$$

Solution.
We use the substitution method to solve this recurrence. We guess as a solution $T(n) = O(n\lg^2 n)$ ie $T(n) \leq cn\lg^2 n$, for some positive constant $c > 0$ and for all $n \geq n_0$. We use induction to prove that this is indeed a solution to the recurrence.

$$\begin{aligned} T(n) &= T(n/4) + T(3n/4) + n\lg n \\ &\leq (cn/4)\lg^2(n/2) + (3cn/4)\lg^2(3n/2) \\ &= cn\lg^2 n + cn(4 + 3\lg^2 3 - 3\lg 3) + cn\lg n(-4 + 3\lg 3/2) \\ &\leq cn\lg^2 n + 1.13cn - 1.6225cn\lg n \\ &\leq cn\lg^2 n + 2cn - 2cn\lg n \\ &\leq cn\lg^2 n \end{aligned}$$

for any $n \geq 2$ and $c > 0$. We choose for boundary condition $T(2) = 2$ so that $T(n) \leq cn\lg^2 n$ would result in $2 \leq c2\lg^2 2 = 2c$, ie $c \geq 1$. Therefore for any $n \geq 2$ and any $c \geq 1$, we have $T(n) \leq cn\lg^2 n$. ■

Exercise 507.

Solve the following recurrence. Use asymptotic notation to express your solution.

$$T(n) = 5T(n^{1/5}) + \log_5 n,$$

where $T(5) = 1$ and $n = 5^{5^k}$ for some k.

Solution.
As $n = 5^{5^k}$, $\log_5 n$ is well defined as it is an integer. We set $n = 5^l$ ie $l = \log_5 n$. Then, $T(n) = T(5^l)$, and we rename $A(l) = T(5^l) = T(n)$. Note that $A(1) = T(5^1) = T(5) = 1$. Then the original recurrence becomes (note again that $l/5$ is an integer, as l itself is also a power of 5).

$$\begin{aligned} A(l) = T(n) &= 5T(n^{1/5}) + \log_5 n \\ &= 5T(5^{l/5}) + l \\ &= 5A(l/5) + l. \end{aligned}$$

The simplified recurrence has solution $A(l) = \Theta(l \lg l)$ (master method). As $n = 5^l$, then $l = \log_5 n$ and $\lg l = \Theta(\log_5 l) = \Theta(\lg \lg n)$. Therefore $T(n) = A(l) = \Theta(\lg n \lg \lg n)$. ■

Exercise 508.

Solve the following recurrence. Use asymptotic notation to express your solution.

$$T(n) = 2T(n/2) + T(n/4) + n^2$$

Solution.
We first note that $T(n) = 2T(n/2) + T(n/4) + n^2 \geq n^2$, and thus $T(n) = \Omega(n^2)$.
Moreover, since

$$T(n) = 2T(n/2) + T(n/4) + n^2 \leq 3T(n/2) + n^2$$

either through master method (case three) or by using the substitution method we deduce that $T(n) = O(n^2)$. Therefore $T(n) = \Theta(n^2)$. Let's show that $T(n) \leq 3T(n/2) + n^2$ has $T(n) = O(n^2)$ as a solution i.e. $T(n) \leq cn^2$ for all $n \geq n_0$. We choose the base case ourselves $T(1) = 1$ which implies $c \geq 1$. By the inductive hypothesis $T(i) \leq ci^2$ we have for $i = n/2$ that $T(n) \leq 3T(n/2) + n^2 \leq 3c(n/2)^2 + n^2 = 3cn^2/4 + n^2$. The latter is at most cn^2 for $c \geq 4$. Therefore for $c = 4$ and $n_0 = 1$ we have shown $T(n) = O(n^2)$. If we combine this with the $T(n) = \Omega(n^2)$ the result follows. ■

Exercise 509.

Solve the following recurrences. You may assume $T(1) = \Theta(1)$, where necessary. Use asymptotic notation. Justify your answers.

a. $T(n) = 16T(n/2) + n^2$.
b. $T(n) = 9T(n/3) + n^4$.
c. $T(n) = 36T(n/6) + n^2$.
d. $T(n) = T(n/2) + n$.
e. $T(n) = 8T(n/8) + n^2$.
f. $T(n) = 16T(n/2) + n$.
g. $T(n) = 9T(n/3) + n^2$.
h. $T(n) = 4T(n/2) + n^4$.
i. $T(n) = 4T(n/2) + n$.
j. $T(n) = T(n/6) + T(2n/3) + 2n$
k. $T(n) = 4T(n/2) + 3n^2$
l. $T(n) = 32T(n/4) + 3n^2$

Solution.

a. We use the master method. $a = 16 \geq 1$, $b = 2 > 1$, $f(n) = n^2$, and $\log_b a = 4$. Then $n^{\lg_b a} = n^4$. As $f(n) = n^2 = O(n^{\lg_b a - 1}) = O(n^3)$, case one of the master method applies, and thus $T(n) = \Theta(n^4)$.

b. We use the master method. $a = 9 \geq 1$, $b = 3 > 1$, $f(n) = n^4$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n^4 = \Omega(n^{2+2}) = O(n^{\lg_b a + 2})$, for $\varepsilon = 2$, case three of the master method applies, and thus $T(n) = \Theta(n^4)$, as in addition $af(n/b) = 9f(n/3) = 9n^4/81 \leq cn^4 = cf(n)$ for say $c = 1/2$.

c. We use the master method. $a = 36 \geq 1$, $b = 6 > 1$, $f(n) = n^2$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n^2 = \Theta(n^{\lg_b a}) = \Theta(n^2)$, case two of the master method applies, and thus $T(n) = \Theta(n^2 \lg n)$.

d. We use the master method. $a = 1 \geq 1$, $b = 2 > 1$, $f(n) = n$, and $\log_b a = 0$. Then $n^{\lg_b a} = n^0 = 1$. As $f(n) = n = \Omega(n^{0+1}) = O(n^{\lg_b a + 1})$, for $\varepsilon = 1$, case three of the master method applies, and thus $T(n) = \Theta(n)$, as in addition $af(n/b) = 1f(n/2) = n/2 \leq cn = cf(n)$ for say $c = 1/2$.

(e) We use the master method. $a = 8 \geq 1$, $b = 8 > 1$, $f(n) = n^2$, and $\log_b a = 1$. Then $n^{\lg_b a} = n$. As $f(n) = n^2 = \Omega(n^{1+1})$ case three of the master method applies, since also $8f(n/8) = 8n^2/64 \leq 1/2n^2$ for $c = 1/2 < 1$, and thus $T(n) = \Theta(n^2)$.

(f). We use the master method. $a = 16 \geq 1$, $b = 2 > 1$, $f(n) = n$, and $\log_b a = 4$. Then $n^{\lg_b a} = n^4$. As $f(n) = n = O(n^{4-2})$, case one of the master method applies, and thus $T(n) = \Theta(n^4)$.

(g). We use the master method. $a = 9 \geq 1$, $b = 3 > 1$, $f(n) = n^2$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n^2 = \Theta(n^2)$, case two of the master method applies, and thus $T(n) = \Theta(n^2 \lg n)$.

It has $a = 4 \geq 1$, $b = 2 > 1$, $f(n) = n^4$. $n^{\log_b a} = n^{\lg 4} = n^2$. Obviously $n^4 = f(n) = \Omega(n^{\log_b a + \varepsilon}) = \Omega(n^3)$, by choosing $\varepsilon = 1 > 0$. Case 3 of the Master method applies since also $af(n/b) = 4(n/2)^4 = n^4/4 \leq cn^4 = (1/2)n^3$, by choosing $c = 1/2 < 1$, and thus $T(n) = \Theta(n^4)$.

(i) It has $a = 4 \geq 1$, $b = 2 > 1$, $f(n) = n$. $n^{\log_b a} = n^{\lg 4} = n^2$. Obviously $n = f(n) = O(n^{\log_b a - \varepsilon}) = \Omega(n)$, by choosing $\varepsilon = 1 > 0$. Case 1 of the Master method applies and thus $T(n) = \Theta(n^2)$.

(j) We use guess and check (aka the substitution method). Let $T(n) \leq cn$. Then

$$T(n) \leq cn/6 + 2cn/3 + 2n = 5cn/6 + 2n = (5c/6 + 2)n.$$

In order to get $5c/6 + 2$ less than or equal to c it suffices to choose any c such that $c \geq 12$. Then $T(n) = \Theta(n)$ (as $T(n) \geq 2n$ as well).

(k) We use the master method. $a = 4 \geq 1$, $b = 2 > 1$, and $n^{\log_b a} = n^2$. $f(n) = 3n^2$ and positive. Therefore, $f(n) = \Theta(n^{\log_b a})$ and therefore $T(n) = \Theta(n^{\log_b a} \lg n) = \Theta(n^2 \lg n)$.

(l) We use the master method. $b = 4 > 1$, $a = 32 = 2^5 = 4^{5/2} = 4^{2.5} \geq 1$, and $n^{\log_b a} = n^{2.5}$. $f(n) = 3n^2$ and positive. Therefore, $f(n) = O(n^{\log_b a - \varepsilon})$ and therefore $T(n) = \Theta(n^{\log_b a}) = \Theta(n^{2.5})$.

■

Exercise 510.

Solve the following recurrences. You may assume $T(1) = \Theta(1)$, where necessary. Use asymptotic notation. Justify your answers.

a. $T(n) = 4T(n^{1/4}) + \log_4 n$, where $T(4) = 1$ and $n = 4^{4^k}$ for some k.
b. $T(n) = 8n + \sum_{i=1}^{\lg n} T(n/2^i)$.
c. $T(n) = T(n/2) + T(2n/3) + n$.
d. $T(n) = 4T(n/2) + n^2/\lg n$, n a power of 2.
e. $T(n) = 4T(n/3) + n^{1/3}$.
f. $T(n) = n^{3/4}T(n^{1/4}) + n$, where $T(n) \leq c$, for all $n \leq n_0$, for some positive constant n_0, c.
g. $T(n) = 9T(n/3) + n\lg^2 n$
h. $T(n) = 3T(n/9) + n^2$
i. $T(n) = 2T(n/4) + \sqrt{n}$
j. $T(n) = 4T(n/2) + n^2$.
k. $T(n) = 16T(n/2) + n$.
l. $T(n) = 3T(n/9) + n$.

Solution.
(a) As $n = 4^{4^k}$, $\log_4 n$ is well defined as it is an integer. We set $n = 4^l$ ie $l = \log_4 n$. Then, $T(n) = T(4^l)$, and we rename $A(l) = T(4^l) = T(n)$. Note that $A(1) = T(4^1) = T(4) = 1$. Then the original recurrence becomes (note again that $l/4$ is an integer, as l itself is also a power of 4).

$$\begin{aligned} A(l) = T(n) &= 4T(n^{1/4}) + \log_4 n \\ &= 4T(4^{l/4}) + l \\ &= 4A(l/4) + l. \end{aligned}$$

The simplified recurrence has solution $A(l) = \Theta(l \lg l)$. As $n = 4^l$, $l = \log_4 n$ and $\lg l = \Theta(\log_4 l) = \Theta(\lg\lg n)$. Therefore $T(n) = A(l) = \Theta(\lg n \lg\lg n)$.
(b) We write the recurrence as follows

$$T(n) = 8n + T(n/2) + T(n/4) + \ldots + T(n/n) = (4n + T(n/2)) + (4n + T(n/4) + \ldots + T(1))$$

We then substitute $n/2$ for n in the original recurrence to get.

$$T(n/2) = 8n/2 + \sum_{i=1}^{\lg(n/2)} T((n/2)/2^i) = 4n + T((n/2)/2) + T((n/2)/4) + \ldots + T(1) = 4n + T(n/4) + T(n/8) + \ldots + T(1).$$

In the derived expression for $T(n)$ we recognize the second parenthesized term to be $T(n/2)$. Therefore,

$$T(n) = (4n + T(n/2)) + (4n + T(n/4) + \ldots + T(1)) = (4n + T(n/2)) + T(n/2) = 2T(n/2) + 4n$$

This is the well-known "mergesort" recurrence ie $T(n) = \Theta(n \lg n)$.
(c) We first observe that $T(n)$ is such that $T(n) \geq n$. We note that the guess and check (substitution method) does not work for $T(n) \leq cn$. A second choice, $T(n) \leq cn^2$ would work, however, as by the inductive hypothesis $T(n/2) \leq c(n/2)^2$ and $T(2n/3) \leq c(2n/3)^2$ and

$$T(n) \leq cn^2/4 + 4cn^2/9 + n = c(25/36)n^2 + n \leq cn^2$$

for large enough n and $c > 0$. It is not, however, the best possible one. **Partial credit will be given to those who suggested such a solution**. This leads us to try a solution $T(n) \leq cn^\alpha$ where α is a constant such that $1 < \alpha < 2$. Similarly as before,

$$T(n) \leq cn^\alpha((1/2)^\alpha + (2/3)^\alpha) + n \leq cn^\alpha$$

The latter inequality holds asymptotically for large n provided that $(1/2)^\alpha + (2/3)^\alpha < 1$. If this value is for example 0.99, then $0.99cn^\alpha + n \leq cn^\alpha$ asymptotically for large n.

A more accurate solution is to try $T(n) \leq cn^\alpha - \beta n$ (β a constant). A similar lower bound can also be proved. For the choice $T(n) \leq cn^\alpha - 6n$, as long as $(1/2)^\alpha + (2/3)^\alpha) \leq 1$, the recurrence can be verified as by the inductive hypothesis $T(n/2) \leq c(n/2)^\alpha - 3n$ and $T(2n/3) \leq c(2n/3)^\alpha - 4n$ and

$$T(n) \leq cn^\alpha((1/2)^\alpha + (2/3)^\alpha) - 3n - 4n + n \leq cn^\alpha - 6n$$

An α such that $(1/2)^\alpha + (2/3)^\alpha = 1$ is one such that $1.293 < \alpha < 1.294$ (α can be found numerically, but the exact value is of no particular interest; **what matters, is that it is a constant between 1 and 2**). Therefore $T(n) = \Theta(n^{O(1)})$, where the exponent of n is some constant between 1 and 2 that can be found as shown above.
(d)

$$\begin{aligned}
T(n) &= 4T(n/2) + n^2/\lg n \\
&= 4(4T(n/4) + (n^2/4)\lg(n/2)) + n^2/\lg n = 4^2T(n/2^2) + n^2/\lg(n/2) + n^2/\lg n \\
&= 4^iT(n/2^i) + n^2/\lg(n/2^{i-1}) + \ldots + n^2/\lg n = 4^iT(n/2^i) + \sum_{j=0}^{i-1} n^2/\lg(n/2^j) = 4^iT(n/2^i) + \sum_{j=0}^{i-1} n^2/(\lg n - j) \\
&= 4^iT(n/2^i) + n^2\sum_{j=0}^{i-1} 1/(\lg n - j)
\end{aligned}$$

We then set $n/2^i = 1$, ie $i = \lg n$. We then get.

$$T(n) = 4^iT(1) + n^2\sum_{j=0}^{i-1} 1/(\lg n - j) = n^2 + n^2\sum_{j=0}^{\lg n - 1} 1/(\lg n - j) = n^2 + n^2\sum_{k=1}^{\lg n} 1/k$$

The sum, from Chapter 3 (equation (3.5)) is equal to $\Theta(\lg\lg n) + \Theta(1)$. Therefore, $T(n) = \Theta(n\lg\lg n)$.
(e) We use the master method. $a = 4 \geq 1$. $b = 3 > 1$. $f(n) = n^{1/3}$ and $n^{\log_b a} = n^{\log_3 4} \approx n^{1.26}$. From case 1 of the master theorem, $f(n) = O(n^{\log_3 4 - \varepsilon})$, by choosing any $0 < \varepsilon < \log_3 4 - 1/3 \approx 0.9$. Therefore $T(n) = \Theta(n^{\log_3 4})$.
(f) We use the iteration method.

$$\begin{aligned}
T(n) &= n + n^{3/4}T(n^{1/4}) \\
&= n + n^{3/4}(n^{1/4} + (n^{1/4})^{3/4}T((n^{1/4})^{1/4})) = n + n^{3/4}(n^{1/4} + n^{3/16}T(n^{1/16})) = n + n + n^{15/16}T(n^{1/16}) \\
&= n + n + n^{1-1/16}T(n^{1/16}) \\
&= 2n + n^{15/16}(n^{1/16} + (n^{1/16})^{3/4}T((n^{1/16})^{1/4})) = 2n + n + n^{63/64}T(n^{1/64}) = 3n + n^{1-(1/4)^3}T(n^{1/4^3}) \\
&= kn + n^{1-(1/4)^k}T(n^{1/4^k})
\end{aligned}$$

We then set $k = \log_4\log_4 n$. Then, $n^{1/4^k} = n^{1/\log_4 n} = 4$. In addition, $n^{1-1/4^k} < n$.

$$T(n) = n\log_4\log_4 n + n^{1-(1/4)^k}T(4) = n\log_4\log_4 n + O(n)T(4) = n\log_4\log_4 n + O(n)$$

Therefore $T(n) = \Theta(n\log_4\log_4 n)$.
g) We use the master method. $a = 9 \geq 1$, $b = 3 > 1$, and $n^{\log_b a} = n^2$. $f(n) = n\lg^2 n$ and positive for $n > 1$. Therefore, $f(n) = O(n^{\log_b a - \varepsilon})$ for $\varepsilon = 1/2$. Then, $T(n) = \Theta(n^{\log_b a}) = \Theta(n^2)$.
h) We use the master method. $a = 3 \geq 1$, $b = 9 > 1$, and $n^{\log_b a} = n^{1/2}$. $f(n) = n^2$ and positive. Therefore, $f(n) = \Omega(n^{\log_b a + \varepsilon})$ for $\varepsilon = 1/2$. In addition $af(n/b) \leq cf(n)$ for $c = 3/81 < 1$. Then, $T(n) = \Theta(f(n)) = \Theta(n^2)$.
i) We use the master method. $b = 4 > 1$, $a = 2 \geq 1$, and $n^{\log_b a} = n^{1/2}$. $f(n) = n^{1/2}$ and positive. Then, $f(n) = \Theta(n^{\log_b a})$ for all n. Therefore $T(n) = \Theta(n^{\log_b a}\lg n) = \Theta(\sqrt{n}\lg n)$.

(j) We use the master method. $a = 4 \geq 1$, $b = 2 > 1$, $f(n) = n^2$, and $\log_b a = 2$. Then $n^{\lg_b a} = n^2$. As $f(n) = n^2 = \Theta(n^{\lg_b a}) = \Theta(n^2)$, case two of the master method applies, and thus $T(n) = \Theta(n^2 \lg n)$.
(k) We use the master method. $a = 16 \geq 1$, $b = 2 > 1$, $f(n) = n$, and $\log_b a = 4$. Then $n^{\lg_b a} = n^4$. As $f(n) = n = O(n^{4-1}) = O(n^{\lg_b a - 1})$, for $\varepsilon = 1$, case one of the master method applies, and thus $T(n) = \Theta(n^4)$.
(l) We use the master method. $a = 3 \geq 1$, $b = 9 > 1$, $f(n) = n$, and $\log_b a = 1/2$. Then $n^{\lg_b a} = n^{1/2}$. As $f(n) = n = \Omega(n^{1/2+1/2}) = O(n^{\lg_b a + 1/2})$, for $\varepsilon = 1/2$, case three of the master method applies, and thus $T(n) = \Theta(n)$, as in addition $af(n/b) = 3f(n/9) = 3n/9 < cn = cf(n)$ for any choice of c such that $1/3 < c < 1$ (e.g. $c = 1/2$). ■

Exercise 511.

Solve the following recurrence. Make your bounds as tight as possible. Use asymptotic notation to express your solution.

$T(n) = T(n/5) + T(3n/4) + 8n$, where $T(1) = 100$.

Solution.

Since $T(n) = T(n/5) + T(3n/4) + 8n$ we have for $n = 1$ from the base case that $T(1) = 100$. We show that $T(n) = \Omega(n)$. By the recurrence we have that $T(n) \geq 8n$ for all $n > 1$ since $T(n/5), T(3n/4)$ are non-negative and the other term is $8n$. In addition, $T(1) = 100 \geq 8 \cdot 1 = 8$. Therefore $T(n) \geq 8n$ for $n = 1$ as well. Therefore $T(n) = \Omega(n)$.

We now prove that $T(n) = O(n)$ or equivalently that there exist positive constants n_2, c_2 such that $T(n) \leq c_2 n$ for all $n \geq n_2$. Proof is by induction on n. **Base Case.** $T(1) = 100 \leq c_2 \cdot 1$ provided that $c_2 \geq 100$. This completes the base case.

Inductive Step. Let $T(i) \leq c_2 i$ be true for all $1 \leq i < n$. We want to show that $T(n) \leq c_2 n$. Since $T(i) \leq c_2 i$ for all $i < n$, it is also true for $i = n/5$ and $i = 3n/4$ as well, and thus by the inductive hypothesis we obtain $T(n/5) \leq c_2 n/5$ and $T(3n/4) \leq 3c_2 n/4$. Then we use these two inequalities in the recurrence to obtain.

$$\begin{aligned} T(n) &= T(n/5) + T(3n/4) + 8n \\ &\leq c_2 n/5 + c_2 3n/4 + 8n \\ &= 19c_2 n/20 + 8n \end{aligned}$$

In order to complete the inductive step we need to prove that $T(n) \leq c_2 n$ or equivalently $19c_2 n/20 + 8n \leq c_2 n$. The latter is equivalent to $8n \leq c_2 n/20$ or $c_2 \geq 160$. Therefore the inductive step is completed and $T(n) \leq c_2 n$ for any $c_2 \geq 160$.

In conclusion, combining the inductive step and the base case we obtain that $T(n) \leq c_2 n$ for all $n \geq 1$ as long as $c_2 \geq 160$. We can fix $c_2 = 160$ and thus show that $T(n) = O(n)$. Therefore since $T(n) = O(n)$ and $T(n) = \Omega(n)$ we obtain that $T(n) = \Theta(n)$. ■

Exercise 512.

Solve the following recurrence. Make your bounds as tight as possible. Use asymptotic notation to express your solution for each one of them.

$T(n) = T(n/4) + T(2n/3) + 8n$, where $T(1) = 40$.

Solution.

Since $T(n) = T(n/4) + T(2n/3) + 8n$ we have for $n = 1$ from the base case that $T(1) = 40$. We show that $T(n) = \Omega(n)$. By the recurrence we have that $T(n) \geq 8n$ for all $n > 1$ since $T(n/4), T(2n/3)$ are non-negative and the other term is $8n$. In addition, $T(1) = 40 \geq 8 \cdot 1 = 8$. Therefore $T(n) \geq 8n$ for $n = 1$ as well. Therefore $T(n) = \Omega(n)$.

(b) We now prove that $T(n) = O(n)$ or equivalently that there exist positive constants n_2, c_2 such that $T(n) \geq c_2 n$ for all $n \geq n_2$. Proof is by induction on n.

Base Case. $T(1) = 40 \leq c_2 \cdot 1$ provided that $c_2 \geq 40$. This completes the base case.

Inductive Step. Let $T(i) \leq c_2 i$ be true for all $1 \leq i < n$. We want to show that $T(n) \leq c_2 n$. Since $T(i) \leq c_2 i$ for all $i < n$, it is also true for $i = n/4$ and $i = 2n/3$ as well, and thus by the inductive hypothesis we obtain $T(n/4) \leq c_2 n/4$ and $T(2n/3) \leq 2c_2 n/3$. Then we use these two inequalities in the recurrence to obtain.

$$\begin{aligned} T(n) &= T(n/4) + T(2n/3) + 8n \\ &\leq c_2 n/4 + c_2 2n/3 + 8n \\ &= 11c_2 n/12 + 8n \end{aligned}$$

In order to complete the inductive step we need to prove that $T(n) \leq c_2 n$ or equivalently $11c_2 n/12 + 8n \leq c_2 n$. The latter is equivalent to $8n \leq c_2 n/12$ or $c_2 \geq 96$. Therfore the inductive step is completed and $T(n) \leq c_2 n$ for any $c_2 \geq 96$. In conclusion, combining the inductive step and the base case we obtain that $T(n) \leq c_2 n$ for all $n \geq 1$ as long as $c_2 \geq 96$. We can fix $c_2 = 96$ and thus show that $T(n) = O(n)$. Therefore since $T(n) = O(n)$ and $T(n) = \Omega(n)$ we obtain that $T(n) = \Theta(n)$. ■

Exercise 513.

Solve the following recurrence. Make your bounds as tight as possible. Use asymptotic notation to express your solution for each one of them.

$T(n) = T(n/8) + T(7n/8) + n.$

Solution.

If we try $T(n) \leq cn$, we realize this guessed solution does not work. One reason is that $n/8 + 7n/8 = n$; if the sum were dn, where d is a positive constant STRICTLY less than 1, our guess would have worked. We then try a solution $T(n) = O(n \lg n)$, i.e. we assume that there exist positive constants c, n_0 such that $T(n) \leq cn \lg n$ for all $n \geq n_0$. We choose $T(1) = 0$. Then $0 = T(1) \leq c \cdot 1 \cdot 0$, i.e. basis of induction is true. We assume that $T(i) \leq ci \lg i$ for all $i < n$ (inductive hypothesis). We use the inductive hypothesis to bound $T(n/8)$ and $T(7n/8)$, and by way of the recurrence we obtain.

$$\begin{aligned} T(n) &= T(n/8) + T(7n/8) + n \\ &\leq c(n/8)\lg(n/8) + c(7n/8)\lg(7n/8) + n \\ &= cn\lg n + (cn/8)(7\lg(7/8) + \lg(1/8)) + n \\ &\leq cn\lg n + (cn/8)(7\lg 1 + (-3)) + n \\ &\leq cn\lg n - 3cn/8 + n \\ &\leq cn\lg n \end{aligned}$$

provided that $c \geq 8/3$. Therefore $T(n) = O(n \lg n)$. We then prove that $T(n) = \Omega(n \lg n)$ in a similar way. He need to use the following result. The entropy function $H(p)$, is defined for any $0 < p < 1$ as follows

$$H(p) = p\lg p + (1-p)\lg(1-p)$$

It attains its minimum for $p = 1/2$, i.e. $H(p) \geq -1$. We show that for some positive constant c, $T(n) \geq cn \lg n$ for asymptotically large n.

$$\begin{aligned} T(n) &= T(n/8) + T(7n/8) + n \\ &\geq c(n/8)\lg(n/8) + c(7n/8)\lg(7n/8) + n \\ &= cn\lg n + (cn)((7/8)\lg(7/8) + (1/8)\lg(1/8)) + n \\ &\geq cn\lg n + (cn)(-1) + n \\ &\geq cn\lg n - cn + n \\ &\geq cn\lg n \end{aligned}$$

provided that $c \leq 1$ (e.g. $c = 1/2$ is such a choice). Therefore $T(n) = \Omega(n \lg n)$. As a conclusion we proved that $T(n) = \Theta(n \lg n)$, where $c_1 = 1/2$, $c_2 = 8/3$ and $n_0 = 1$. ■

7.5 Recurrences and algorithm analysis: warming up

Exercise 514.

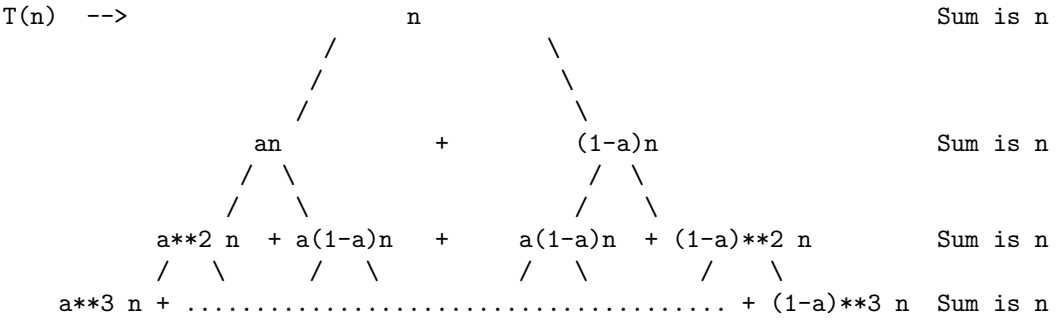
Consider the following variation of MergeSort: Instead of an even split, we split 10-percent on one-size and 90-percent on the other. Then the recurrence for the running time becomes

$$T(n) = T(0.1n) + T(0.9n) + \Theta(n)$$

In general we might have an an split and (1-a)n$splitwhere 0 < a \leq 1/2$. The recurrence then becomes after simplifying the last term (one may use cn instead for some constant $c > 0$).

$$T(n) = T(an) + T((1-a)n) + n$$

Solution. **A solution using the Recursion tree method.** Before you go back to Problem 12 of this Problem Set, note the catch there with $a + b < 1$ but in our case $a + (1-a) = 1$! This recurrence is easier solved by the recursion tree approach (used in MergeSort and Fibonacci sequence in class). A very flat tree is shown in textual/graphic form below. Isolate the leftmost and the rightmost paths of this tree.
The leftmost path is $n, an, a^2n, a^3n, \ldots$ and the rightmost path is $n, (1-a)n, (1-a)^2n, (1-a)^3n, \ldots$.
How long is the leftmost path? It is L long if $a^L n = 1$. Taking logarithms base two $L \lg a + \lg n = 0$, i.e. $L = -\lg n / \lg a$. Substitute 0.1 for a and we get the desired answer.
How long is the rightmost path? It is R long if $(1-a)^R n = 1$. Taking logarithms base two $R \lg(1-a) + \lg n = 0$, i.e. $R = -\lg n / \lg(1-a)$. Subsitute 0.1 for a or 0.9 for $1-a$ and we get the desired answer.



Remark. As a byproduct of this problem, we conclude that in the context also of the QuickSort algorithm of Subject 6, if there is a PARTITION function (we shall call the implementation of the designated textbook GTP) that always causes a 10-percent and 90-percent split or in the more general case an an-$(1-a)n$ split, then the running time of quicksort that uses this partition is going to be $O(n \lg n)$. This is because we just proved that the depth of the recursion is $O(\lg n)$ and at each level, partition spends total time $O(n)$.

A solution using the Substitution method. We could use the substitution method to solve this recurrence.
(1.) Ω part: $T(n) = \Omega(n \lg n)$ uses a line of argument similar to the O part to follow.
(2.) O part: We show $T(n) = O(n \lg n)$ i.e. there exist positive constant c, n_0 such that $T(n) \leq cn \lg n$, for all $n \geq n_0$ by utilizing strong induction. Since $an < n$ and $(1-a)n < n$ the induction hypothesis applies to $T(an)$ and $T((1-a)n$.

$$\begin{aligned} T(n) &= T(an) + T((1-a)n) + n \\ &\leq can \lg(an) + c(1-a)n \lg((1-a)n) + n \\ &= cn \lg n + cn(a \lg a + (1-a) \lg(1-a)) + n \leq cn \lg n - cnA + n \\ &\leq cn \lg n + (1 - c_2 A)n \leq cn \lg n \qquad \text{for} \quad c > 1/A. \end{aligned}$$

where we used that the entropy function i.e. the expression $a \lg a + (1-a) \lg(1-a)$ is negative for $0 < a < 1$. Let that negative value be called $-A = a \lg a + (1-a) \lg(1-a)$, where $A > 0$. We obtain the final answer (i.e. inductive step) by making sure that the term $(1 - cA)n$ is negative which is so for $1 - cA < 0$ i.e. $c > 1/A$. Note

that A is constant because a is constant. **If a were not a constant, then A would not have been a constant either, and c would not have been bounded by a constant either.**

■

Exercise 515.
Solve the following divide and conquer recurrence relation exactly for the case where n is a power of two

$$S(n) = S(n/2) + 3n, \qquad S(1) = 1$$

Solution.
Let $n = 2^k$. Then from $S(n) = S(n/2) + 3n$, $S(1) = 1$ we get:

$$\begin{aligned} S(n) = S(2^k) &= S(2^{k-1}) + 32^k \\ &= S(2^{k-2}) + 32^{k-1} + 32^k \\ &= S(2^{k-3}) + 32^{k-2} + 32^{k-1} + 32^k \\ &= \vdots \\ &= S(2^0) + 3(2^1 + 2^2 + \ldots 2^k) \\ &= 1 + 3(2^1 + \ldots + 2^k) \\ &= 1 + 3 \cdot 2 \cdot (2^k - 1) = 6 \cdot 2^k - 5 = 6n - 5 \end{aligned}$$

■

Part IV

Graphs

Chapter 8

Introduction to graph theory

8.1 Short review on directed and undirected graphs

Definition 8.1 **Graph G**

A Graph $G = (V,E)$ is represented by its sets of vertices V and its set of edges $E \subseteq V \times V$. We may also write $V = V(G)$ and $E = E(G)$ respectively.

Definition 8.2 **Subgraph G**

For a subgraph G_1 of G we have $G_1 = (V_1,E_1)$, where $V_1 \subseteq V$ and $E_1 \subseteq E$. We might also write $V_1 = V(G_1)$ or $E_1 = E(G_1)$ for V_1 and E_1.

Definition 8.3 **Edges of an undirected graph G**

Let $e \in E = E(G)$ for a graph G. For an undirected graph $e = (u,v)$ implies no direction. Thus one would mean $e = \{u,v\}$ as (u,v) in set theory is the ordered pair of u and v. The edge e is incident on u and v.

Definition 8.4 **Edges of an directed graph G**

Let $e \in E = E(G)$ for a graph G. For an directed graph $e = (u,v)$ implies a direction from u to v. Thus one could not write $e = \{u,v\}$ as $\{u,v\}$ in set theory is the set containing elements u and v in no particular order. The edge e is incident from u to v.

The adjacency preposition used is thus "on" for an undirected graph; for a directed graph we use "from" and "to" instead, as needed.

8.1.1 Degree of a vertex

Definition 8.5 **Degree of u in undirected G**

The degree of a vertex in an undirected graph is the number of edges incident ON that vertex.

Definition 8.6 **Degree of u in directed G**

The degree of a vertex in a directed graph is the number of edges incident FROM or TO that vertex.

Definition 8.7 **In-Degree of u in directed G**

The indegree (in-degree) of a vertex in a directed graph is the number of edges incident TO that vertex.

Definition 8.8 **Out-Degree of u in directed G**

The outdegree (out-degree) of a vertex in a directed graph is the number of edges incident FROM that vertex.

Theorem 8.1

Let $d(u)$, $i(u)$, $o(u)$ be the degree, indegree and outdegree of a vertex. Then

$$\sum_{u\in V} d(u) = 2m, \qquad \sum_{u\in V} i(u) = m, \qquad \sum_{u\in V} o(u) = m.$$

A path implies that edges of the path are distinct and vertices of the path are also distinct.
A (sometimes used) trail implies that edges are distinct but not vertices.
A walk implies that neither edges need to be distinct nor vertices need to be distinct.

8.1.2 Simple graphs

Several times the term simple applies to undirected graphs only. A graph is simple if it has no self-loops and no multiple edges.
For directed graphs simple requires no multiple edges but self-loops might be okay.
By default a graph would be simple at the extreme (no self-loops, no multiple edges, directed or undirected).

8.1.3 Regular graphs

An undirected graph is regular of degree k if every vertex has degree k.

8.1.4 Cycles, closed trails, tours, circuits in graphs

A cycle for an undirected graph is of length at least three as it is by default simple.
For a directed graph a cycle can be of length two as there can be an edge from u to v and another edge from v to u.
A cycle is thus a path where an extra edge is implied: it connects the last vertex of the path to the first vertex in a wrap-around fashion.
A closed trail is sometimes defined as a tour or a circuit. In a circuit not all edges need to appear in it; in a tour all edges of G appear in it.
The term Euler tour (or Eulerian tour) will be used later and will imply exactly this. But note that the term Euler circuit might also be used as well.

8.1.5 Connectivity and subgraphs

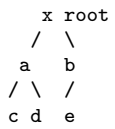
Definition 8.9

A connected undirected graph $G = (V, E)$ is a tree T if it is acyclic (no cycles). If it is not connected, then it is called a forest. A spanning tree of G is a tree that touches all of its vertices.

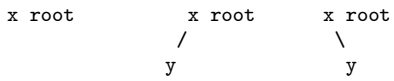
Let $G = (V, E)$. For the subgraph T of G, we have $T = (V, E_1)$, where $E_1 \subseteq E$. We might also write $V = V(T)$, or $E_1 = E(T)$ for V and E_1) respectively.

8.1.6 Trees

For a graph, we use the term vertex to denotes elements of V. For a tree we use the term node instead. The latter use requires some caution. For a rooted tree (where a vertex i.e.node is indicated as the root of the tree, we can define the degree of a node to be the number of its children. Thus for the example below, the degree of node a is two and its two children are c and d. For vertex a though (of the graph that the tree is), the degree of a is the number of edges incident on it and thus it is three instead!



A binary tree is a rooted tree. For every node its children are also ordered and thus there exists a left and right ordering of its children. An empty (implied rooted and ordered) tree is a binary tree. Otherwise it has a designated node as its root; the root has a left-subtree that is (recursively) a binary tree and a right-subtree that is a binary tree as well.
All three trees below are binary trees.



8.1.7 Hamiltonian paths and cycles

Definition 8.10 **Hamiltonian cycle**

A Hamiltonian cycle is a (simple) cycle that contains all the vertices of the graph.

Definition 8.11 **Hamiltonian path**

A Hamiltonian path is a path that contains all the vertices of the graph.

8.1.8 Matching in bipartite graphs

Definition 8.12 **Bipartite graphs**

A bipartite (undirected) graph $G = (V,E) = (A \cup B, E)$, where $V = A \cup B$ and $A \cap B = \emptyset$, is a graph where an edge $e = (a,b)$ is incident on a vertex $a \in A$ and on a vertex $b \in B$.

Thus edges of a bipartite graph are from A to B (or from B to A for that matter).

Definition 8.13 **Matchings in bipartite graphs**

For a bipartite graph $G = (V,E) = (A \cup B, E)$, a matching M is is a subset of E such that for every vertex of A there is exactly one edge of M that is incident on it and the other end point of the edge is in B, and no vertex b of B belongs to more than one edge in M.

In other words a matching in a bipartite graph is a set of edges without common vertices i.e. no two edges are incident on the same vertex.

8.1.9 Matchings in Graphs.

The definitions can extend to an arbitrary graph.

8.1.10 Saturating vertices and sets

Definition 8.14 **Saturating vertex**

A vertex u is saturated or is a saturating vertex if edge (u,v) belongs to a matching M.

Saturation is with respect to a matching and thus u is a saturating vertex with respect to M if $u \in S(M)$.

Definition 8.15 **Saturating set S of an edge**

For a graph $G=(V,E)$ and an edge of it $e=(a,b)\in E$, with $a,b\in V$ we define $S(e)=\{a,b\}$.

Definition 8.16 **Saturating set S of a set of edges M**

For a graph $G=(V,E)$ and an edge set M of G i.e. $M\subseteq E$, we define $S(M)$ as follows

$$S(M)=\cup_{e\in M}S(e).$$

Definition 8.17 **Saturating set S of matching M**

For a matching M the vertices of the matching belong to the saturating by the matching vertex set $S(M)$.

$$S(M)=\{z|(z,b)\in M,b\in B,\ \textbf{OR}\ (a,z)\in M,a\in A\}.$$

8.1.11 Edge cover

Definition 8.18 **Edge cover of a graph**

Set C_E is defined to be an edge cover of graph $G=(V,E)$, if $C_E\subseteq E$ and for every $v\in V$

$$\forall v\in V\Rightarrow v\in S(C_E).$$

8.1.12 Vertex covers

Definition 8.19 **Vertex cover of a graph**

Set C is defined to be a vertex cover of graph $G=(V,E)$, if $C\subseteq V$ and for every $e\in E$ the following holds.

$$\forall e\in E\Rightarrow C\cap S(e)\neq\emptyset.$$

Definition 8.20 **Vertex cover C of a graph with respect to an edge set M**

Set C is defined to be a vertex cover of graph $G=(V,E)$ with respect to an edge set M, if $C\subseteq V$, and $M\subseteq E$, and for every $e\in M$ the following holds.

$$\forall e\in M\Rightarrow C_M\cap S(e)\neq\emptyset.$$

8.1.13 Alternating paths

Definition 8.21 **Alternating path**

Given a reference matching M an alternating path relative to M or an M-alternating path is a path that alternates between edges of M and edges not in M.

Definition 8.22 **Alternating paths in bipartite graphs**

Let M be a matching of a bipartite graph $G = (A \cup B, E)$. An alternating path of G relative to M is a set of edges alternating between edges of M end edges in $E - M$ (i.e. not in M).

We can extend this definition for arbitrary graphs as well.

Definition 8.23 **Augmenting path relative to a matching M**

It is a path in fact an alternating path that starts and ends with a vertex that is not saturated by M.

8.1.14 Augmenting paths

Definition 8.24 **Augmenting path**

Given a reference matching M an augmenting path relative to M or an M-augmenting path is an M-alternating path whose end-points are not in M.

8.1.15 Maximum, maximal and complete matchings

Definition 8.25 **Maximum matching in bipartite graphs**

For a bipartite graph $G = (A \cup B, E)$, a matching is maximum if it contains the largest number of edges possible.

Definition 8.26 **Maximum matching in graphs**

For a graph $G = (V, E)$ rather than a bipartite graph, a maximum matching saturates as many vertices of V as possible. Therefore M is maximum if there is no matching M_1 wih $|M| < |M_1|$.

Definition 8.27 **Maximal matching**

A matching M that cannot be made larger by adding one more edge. However it might be augmented by utilizing an augmented path. Therefore M is maximal if there is no matching M_1 wih $M \subset M_1$.

A maximum matching is a matching M of maximum cardinality that saturates as many vertices of A as possible. Thus $|M| \leq |A|$. It is also possible that we have more matchings of the same cardinality $|M|$.

Definition 8.28 **Complete matching in bipartite graphs**

For a bipartite graph $G = (A \cup B, E)$, a complete matching is a matching M of maximum cardinality that saturates all vertices of A. Thus $|M| = |A|$.

Definition 8.29 **Complete matching in graphs**

For a graph rather than a bipartite graph, a complete matching is of size $\lfloor n/2 \rfloor$.

8.1.16 Perfect matchings

Thus if $d_M(u)$ defines the degree of vertex u in the subgraph $G_M = (A \cup B, M)$, we have $d_M(a) = 1$ and $d_M(b) \leq 1$ for all $a \in A$ and all $b \in B$.

Definition 8.30 **Perfect matching in bipartite graphs**

For a bipartite graph as previously defined, a perfect matching saturates all vertices of A and B. (It is thus a complete matching with respect to A and also a complete matching with respect to B).

Therefore an implication is that $|A| = |B|$ for a perfect matching to exist

Definition 8.31 **Perfect matching in graphs**

For a graph rather than a bipartite graph, a perfect matching is of size $n/2$, n is even.

8.1.17 Neighborhood of a vertex set

Definition 8.32 **Neighborhood of a vertex set**

A bipartite graph $G = (A \cup B, E)$, where $V = A \cup B$, and A and B define its bipartition. Let set S be such that $S \subseteq A$. We define by $N(S)$ the vertices of B that are adjacent to at least one vertex of S.

We can define $N(S)$ for a set $S \subseteq B$ likewise.

8.2 Elementary graph theory

Exercise 516.
We define a graph not very formally as follows. $\{\{1,2,3\},\{e_1=(1,2),e_2=(2,3),e_3=(1,2),e_4=(2,2)\}\}$ The assumption is that the graph is an undirected graph.
Do it properly. Is the graph simple? Does it have multiple edges?

Solution.
We can define this graph G formally as $G=(V,E)$.

$$V=\{1,2,3\},$$

and

$$E=\{e_1=(1,2),e_2=(2,3),e_3=(1,2),e_4=(2,2)\}$$

The graph however is not simple: it contains a self loop in the form of edge e_4, and multiple edges e_1 and e_3 between two vertices 1 and 2. The following is a simple graph derived from the earlier described one. $V=\{1,2,3\}$, and $E=\{e_1=(1,2),e_2=(2,3)\}$. In the remainder we will be dealing with the latter graph NEVER with the former one. ■

Exercise 517.
For the simple graph $G=(V,E)$ with $V=\{1,2,3\}$, and $E=\{e_1=(1,2),e_2=(2,3)\}$ assuming that it is undirected, provide information about its properties.

Solution.
This graph is simple, $n=3$ and $m=2$. Degree-wise, $d(1)=1$, $d(2)=2$ and $d(3)=1$. Therefore $\sum_{w\in V} d(w) = d(1)+d(2)+d(3)=1+2+1=4=2\cdot m$. In the remainder, we either write $E=\{(1,2),(2,3)\}$, or $E=\{e_1,e_2\}$, and define separately $e_1=(1,2)$ and $e_2=(2,3)$. The graph is also connected and consists of one connected component. The vertex set of this one connected component is V. G does not have a cycle. ■

Exercise 518.
Let $G=(V,E)$ be a simple directed graph with $V=\{1,2,3\}$, and $E=\{e_1=(1,2),e_2=(2,3),e_3=(2,1)\}$. Provide information about G.

Solution.
It is described as simple which implies e_1,e_3 are not multiple edges but two different edges of a directed graph. Therefore an absence of the word directed should not have caused a confusion. This graph is simple, $n=3$ and $m=3$. Degree-wise, $d(1)=2$, $d(2)=3$ and $d(3)=1$. Therefore $\sum_{w\in V} d(w)=d(1)+d(2)+d(3)=2+3+1=6=2\cdot m$. Furthermore, $\mathbf{i(1)}=1$, $\mathbf{i(2)}=1$, $\mathbf{i(3)}=1$, and $\sum_{w\in V} i(w)=3=m$. Finally, $\mathbf{o(1)}=1$, $\mathbf{o(2)}=2$, $\mathbf{o(3)}=0$, and $\sum_{w\in V} o(w)=3=m$. From now on, every introduced graph will be explicitly referred to as undirected or directed. ■

Exercise 519.
For a (simple) undirected graph $G = (V, E)$ with $|V| = n$ and $|E| = m$, the sum of the degrees of all the vertices of the graph is an even number. Moreover this sum is $2m$.Thus the following is true.

$$\sum_{w \in V} d(w) = 2m$$

Solution.
Consider edge $(u, v) \in E$ of G. The edge contributes one to the degree $d(u)$ of u and one to the degree $d(v)$ of v i.e. a total of 2 to the sum $\sum_{w \in V} d(w)$. There are m edges in E, and thus the result follows immediately.

■

Exercise 520.
For a (simple) directed graph $G = (V, E)$ with $|V| = n$ and $|E| = m$, the sum of the degrees of all the vertices of the graph is an even number. Moreover this sum is $2m$.Thus the following is true.

$$\sum_{w \in V} d(w) = 2m$$

Furthermore,

$$\sum_{w \in V} i(w) = m = \sum_{w \in V} o(w)$$

Solution.
Every edge $e = (u, v)$ is counted twice once to the degree of one end-point i.e. $d(u)$ and once to the degree of the other end-point i.e. $d(v)$. Thus

$$\sum_{w \in V} d(w) = 2m$$

The rest follows immediately. ■

Exercise 521.
For an undirected graph $G = (V, E)$, the number of vertices of odd degree is an even number.
Examine a similar result for a directed graph $G = (V, E)$.

Solution.
If the number of vertices of odd degree were odd, then their contribution to the sum $\sum_{w \in V} d(w)$ would have been $\text{odd} \times \text{odd}$ which is an odd number. The contribution of the even degree vertices, whether the number of such vertices is odd or even would be an even number.
Adding the contributions of odd-degree and even-degree vertices is equivalent to adding an odd number and an even number. The result is an odd number. We do know from a prior theorem that this result $2m$ i.e. an even number. Thus the number of vertices of odd degree MUST be even! ■

Exercise 522.

In an undirected graph $G = (V, E)$ with at least two vertices we have at least two vertices with the same degree. (Reminder: graph is simple i.e. no self-loops, no multiple vertices.)

Solution.
We prove the result by induction.
Base case. If we have two vertices and they are isolated, their degree is 0. If there is an edge connecting them the degree of both vertices is one. Base case completed.
Inductive step. Let the result be true for all graphs with n or fewer vertices.
Suppose we have a graph G with $n+1$ vertices.
Case 1. One vertex say u has degree 0 i.e. it is an isolated vertex. Then the remainder $G-u$ has n vertices. By induction there are two vertices say a, b with the same degree. Adding to $G-u$ the isolated vertex u would not changes the degrees of a, b. Result proven.
Case 2. All vertices have degrees greater than zero i.e. $1 \leq d(u) < n$. There are n vertices and $n-1$ degrees the values $1, 2, \ldots, n-1$. By the pigeonhole principle there are must be two vertices of the same degree (value). Result proven and proof completed.

■

Exercise 523.
Find the number of vertices in CBT (Complete Binary Tree) of height h. Then the number of internal nodes and the number of external nodes (leaves).

Solution.
It is

$$1+2+2^2+\ldots+2^h=2^{h+1}-1.$$

The number of internal nodes in a CBT of height h is 2^h-1. The number of external nodes in a CBT of height h is 2^h.

■

Exercise 524.
A complete binary tree has n nodes. How many of them are leaves, and how many of them are internal nodes ? What is its height ?

Solution.
If x is the number of internal nodes then $x+1$ is the number of external nodes and thus $2x+1=n$. Other details omitted. ■

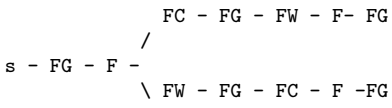
Exercise 525.

A wolf, a goat, and a cabbage are on one bank of a river. A boatman will take them across, but can take only one at a time. The wolf and goat cannot be left together on either bank of the river without the boatman, nor can the goat and the cabbage. Using a graph model determine the shortest time required for the boatman to achieve his task.

Solution.

Let the W(olf), G(oat), and C(abbage) be initially on the left bank of the river. We must show that the B(oatman) can take them to the right bank without any losses and in the shortest time possible. We use the following graph representation of the problem. The vertices below correspond to legal occurrences of W, C, G, B on the left bank of the river, and our aim is to find a shortest path between vertex $WGCB$ (all on the left bank) to $\oslash$ (none on the left blank). The edges of the graph correspond to legal moves under which the boatman can take some of the animals from(to) the left bank to(from) the right without causing any losses of animals on either bank.

Let the W(olf), G(oat), C(abbage), and Boatman be initially on the left bank of the river. We use the following graph representation of the problem. The vertices below correspond to legal occurrences of W, C, G, B on the left bank of the river, and our aim is to find a shortest path between vertex $WGCB$ (all on the left bank) to $\emptyset$ (none on the left blank). The edges of the graph correspond to legal moves under which the boatman can take some passenger from (to) the left bank to (from) the right without attrition on either bank.



■

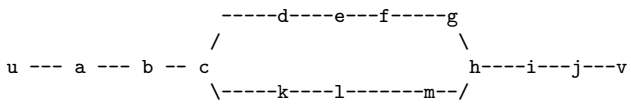
8.3 Trees, forests and connectivity in undirected graphs

Exercise 526.

The following definitions are for undirected graphs. A **forest** is an acyclic undirected graph. A **(free) tree** or just a **tree** is a connected acyclic undirected graph. A connected forest is thus a tree.
Let G be an undirected graph. The following statements are equivalent.

1. G is a free tree.
2. G has exactly one path between any pair of vertices.
3. G is connected but if any edge $e \in E(G)$ is removed from E the resulting graph is disconnected.
4. G is connected and $|E| = |V| - 1$.
5. G is acyclic and $|E| = |V| - 1$.
6. G is acyclic but if any edge is added to E the resulting graph contains a cycle.

Solution.
From (1) to (2) and from (2) to (1). We show how 1 leads to 2 and the other way around.
Let G be a tree and u, v are two of its vertices. Since G is a tree there is a simple path from u to v. If there were to be two paths from u to v, find the first vertex that the paths deviate starting from u. Do the same moving backwards from v. (It is possible that the two paths deviate immediately after u or just before reaching to v.) Then one would define a cycle from the point of convergence prior to deviation following the first path to the point of next convergence and then through the second path back to the point of original convergence. This defines a cycle contradicting to the acyclicity of G



From 2 to 1: Let G have a unique path from u to v for every u, v: this implies that G is connected. To prove G is a tree (a.k.a. free tree) it suffices to prove it is acyclic. If G were to have a cycle, and a, b are two vertices of the cycle, then two paths from a to b would have been defined contradicting the existence of unique paths from any u to any v.
From (1) to (3). Let G be a tree. If $e = (u, v) \in E(G)$ is an edge of G, it creates a path from u to v. By part (2) this path is unique. Thus its removal from $E(G)$ and G causes no path in $G - e$ from u to v. Thus $G - e$ becomes disconnected.
From (3) to (1). Show that G is a tree: G is given to be connected with the property that the removal of any edge from it makes it disconnected. We just nee to show that G is acyclic. If it had a cycle let $e = (u, v)$ be an edge of that cycle. In the cycle there are two paths from u to v: one through the edge $e = (a, b)$ and one the other way around. But we are told that the removal of e should leave G disconnected which would not have been the case if G had been disconnected. Thus G must be acyclic and therefore a tree.
From (1) to (4) and (5). If G is a (free) tree it is both connected and acyclic. We show $|E| = |V| - 1 \Rightarrow m = n - 1$ by induction as follows.
For $n = 1$ and $n = 2$ the result is true by inspection since $m = 0$ and $m = 1$ respectively. A tree with one vertex ($n = 1$) has no edges and a tree with two vertices has one edge, the one connecting the two vertices. Thus number of edges is one less than the number of vertices for $n = 0$ but also for $n = 1$. Let the claim be true for trees with $n - 1$ vertices or less. Consider a tree with n vertices. Pick edge $e = (a, b)$ and remove it from the tree. The graph (tree) G is split into a number (two) of connected components. This is because otherwise there would still be a path from a to b and adding to that path e the tree would have had a cycle. Let the two connected components be T_1 and T_2 with n_1, n_2 vertices respectively where $n_1 + n_2 = n$. Each one has $n_1 < n$ and $n_2 < n$ and are both connected and acyclic inheriting the properties of T thus by induction $m_1 = n_1 - 1$ and $m_2 = n_2 - 1$. The $m = m_1 + m_2 + |\{e\}| = n_1 - 1 + n_2 - 1 + 1 = n - 1$. Both (4) and (5) have been shown.
(An alternative inductive step. Pick a path P of maximum length in tree. The first vertex or the last vertex has

degree one. Call it u and the edge incident on it e. (If its degree were two, it would lead either to a longer path, or to a previously traverse vertex indicating a cycle in a tree, an impossibility). Then $G-u$ has one fewer vertex and one fewer edge as e the edge leading to u is not in $G-u$. By induction $G-u$ is a tree of course and has $n-1$ vertices and $n-2$ edges. Adding u and e we obtain G has n vertices and $n-1$ edges.

From (4) to (1). G is connected and has n vertices and $n-1$ edges. If it is a tree we are done otherwise it is not a tree and thus it has a cycle. Remove an edge from that cycle. The remainder either has not a cycle (it is a tree) or still has a cycle; continue likewise until it becomes a tree (and is acyclic). Then the end-result must have by (4), $n-1$ edges and thus the original graph had n or $n+1$ or more edges as implied by the deconstruction. This is a contradiction to the original graph having $n-1$ edges.

From (1) to (6) and (6) to (1). Use (3). (Or to go from (6) to (1) argue that T must be connected and thus it is a free tree. If T was not connected, there would be two or more connected components. Adding an edge from one to the other would not cause a cycle contradicting to (6) though!) ■

Exercise 527.

The following claims apply to undirected graphs. Some of the properties are for an undirected graph that is also a tree: an acyclic connected undirected graph. If the undirected graph is acyclic and not connected, then it is a forest.
(1) A tree T has exactly one path between any pair of vertices.
(2) A tree T with at least two vertices has at least two leaves (vertex with degree one).
(3) A tree T (undirected connected graph) with n vertices has $n-1$ edges.
(4) Every connected graph has a spanning tree.

Solution.
(1) Let u and v be two distinct vertices of a tree T. Since T is connected there is a path from u to v. Suppose there are two paths from u to v. Removing common edges (and one vertex per removed edge) from those two paths, leads to a cycle. But T is acyclic; thus no two paths exist.
(2) Pick a vertex a. Since graph has at least two vertices and is connected a has a an edge. Let it lead to b, the other end-point of the edge. Either we reached a leaf or b points to c, another vertex through its second edge. Eventually we reach say z. Vertex z has degree at least one. If z is not a leaf either we can go on, or hit the path again leading to a cycle that is not supposed to exist in T. Thus z is a leaf. Note that if a is not a leaf we can extend the path from a to z, by expanding it through vertex a's other (second) edge. Eventually we are going to hit a leaf as well by the same argument. An alternative argument follows. Since tree T has n vertices with $n \geq 2$, all paths in T are of length no more than $n-1$. In it there is a longest path $u_1, u_2, \ldots, u_k$. Both u_1 and u_k must be of degree 1. Otherwise the path can be extended by one more edge say to v, w respectively, an additional neighbor of u_1, u_k respectively. In the remote case v, w are one of $u_2, \ldots, u_{k-1}$, we have discovered a cycle that cannot exist since T is a tree i.e. it is an acyclic graph. (There is another proof, see a later problem, that uses degree arguments.)
(3) By induction. We could start induction for $n = 1$ vertex and $n - 1 = 0$ i.e. no edges. A tree with two vertices has one edge the one connecting the two vertices. Thus number of edges is one less than the number of vertices for $n = 0$ but also for $n = 1$. Let the claim be true for trees with $n-1$ vertices or less. Consider a tree T with n vertices. It must have two leaves by part (2) above. Let u be such a leaf, and v its neighbor. Let T' be created by removing u and edge (u, v). T' has $n-1$ vertices and is acyclic AND connected. This is because T it came from was acyclic and the removal of u retained the acyclicity and connectedness of T'. Thus T' is a tree with $n-1$ vertices. By the induction hypothesis it has $n-2$ edges. Thus T obtained from T' by adding u and edge (u, v) has $(n-2)+1 = n-1$ edges for its $(n-1)+1 = n$ vertices.
(4) By induction on edges. A graph G that is connected with zero edges is a single vertex and it is also a tree T by default. Let any connected graph with at most $m-1$ edges have a spanning tree. Let G have $m \geq 1$ edges. If G is a tree, then it is its own spanning tree. Otherwise it contains a cycle. Remove an edge $e = (u, v)$ from the cycle. The remainder is still connected with same number of vertices and fewer (by one edges). So by induction $G - \{e\}$ has $m-1$ edges and it has a spanning tree which is also a spanning tree for G that has one more edge! (A non-inductive proof follows in a separate problem.)

Exercise 528.

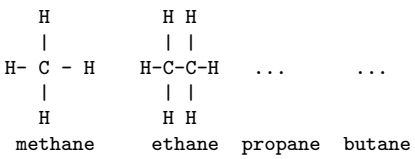
The following are true for undirected graphs $G = (V, E)$. Some of the properties are for an undirected graph that is a tree.
(1) If G is connected, then it has at least $n-1$ edges.
(2) A spanning tree of a connected undirected graph with n vertices has $n-1$ edges, and it always exists.
(3) Let G be a connected undirected graph with n vertices and n edges. G has a cycle.
(4) T is a tree but $T-e$ is disconnected for every edge $e \in E(T)$.
(5) If G has a spanning tree T, then T it is connected.

Solution.
(1) If G is connected it has a spanning tree i.e. at least $n-1$ edges which are edges of G, by part (4) of the previous problem. If G has exactly $n-1$ edges, then G is a tree and the edges are those of its spanning tree.
(2) From part (3) of the previous problem.
(3) G is connected so there is a spanning tree T with n vertices and $n-1$ edges. Let edge $e = (u, v)$ be the one missing from T to bring the edge sum to n edges. In T there is a path p from u to v; the addition to it of e causes a cycle. Moreover the cycle is unique. If there was another cycle call it c, and c does not contain e then c is part of T contradicting to T having no cycles (it is a tree). Otherwise c contains e and another path from u to v. If there are two paths from u to v in T this contradicts to the treeness of T. Thus c is a double count, not a second cycle.
(4) If $e = (u, v)$ is an edge of T, it creates a path from u to v. By part (1) of the previous problem, this path is unique. Thus its removal from T causes no path in $T-e$ from u to v. Thus $T-e$ is disconnected.
(5) G has a spanning tree T. Let u and v be in G. Since T is a spanning tree and connected u and v have a path in T and also in G. Repeat for every u and v in G. ■

■

Exercise 529.

A carbon atom has four free connections that can form convalent bonds with other atoms including carbon atoms or hydrogen atoms for example. A paraffin, formally known as alkane(s), is indicated below where the number of carbon atoms is $n = 1$ and the number of hydrogen atoms $2n+2 = 4$. It is also known as methane. The paraffin with $n = 2$ is known as ethane. A paraffin, in general, has the molecular type C_nH_{2n+2}. Show that the molecule of an alkane is a tree.



Solution.

The number of vertices of the graph of an alkane has $3n+2$ vertices; n of them map to Carbon atoms, and $2n+2$ to Hydrogen. Then n Carbon atoms has degrees 4 each and Hydrogen atoms have degree one each. Thus the sum of the degrees is $4n+2n+2 = 6n+2$, and thus the number of edges is $3n+1$, one less than the number of vertices $3n+2$. Since the graph of the atoms is connected it must be a tree. ■

Exercise 530.
If every vertex in a undirected graph has degree at least $(n-1)/2$ then the graph is connected.

Solution. Pick x,y with degree $(n-1)/2$ or more. Then x,y either they are adjacent and we are done. Or they have a common neighbor since their degrees add up to $((n-1)/2+(n-1)/2) = n-1$ and there are only $n-2$ other vertices. One of them must be common to x,y If there is a common neighbor, then x,y are connected through it.

■

Exercise 531.

Show that if an undirected graph $G = (V, E)$ (with no self loops or multiple edges) has more than $\frac{1}{2}(n-1)(n-2)$ edges, then it must be connected.

Solution.

Assume that G is not connected. We can therefore separate V into two non-trivial partitions V_1 and V_2, where $V = V_1 \cup V_2$ and E contains no edges from $V_1 \times V_2$. The maximum possible number of edges in an n vertex graph is $\frac{n(n-1)}{2}$ (each of n vertices is connected to $n-1$ other vertices, so there are $n(n-1)$ end points and half as many edges). But $|V_1||V_2|$ of those potential edges cannot appear in G. Since $|V_1| + |V_2| = n$ and both sets are non-empty, that product is at least $n-1$ (the minimum of $x(n-x)$ on the range $1 \leq x < n$). Thus the number of edges in an n vertex unconnected graph is at most

$$\frac{n(n-1)}{2} - (n-1) = \frac{1}{2}(n-1)(n-2)$$

■

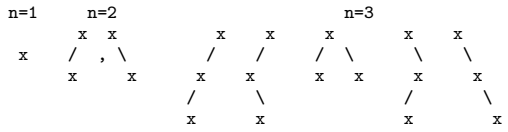
Exercise 532.
Every tree has at least two vertices of degree 1.

Solution.
Let $d(i)$ be the degree of vertex i. We know that $\sum_i d(i) = 2(n-1) = 2n-2$. If $d(i) \geq 2$ for all i, then $\sum_i d(i) \geq 2n > 2n-2$. Thus there are at least two vertices with degree 1.
An alternative argument follows. Let $d(i)$ be the degree of vertex i. We know that $\sum_i d(i) = 2(n-1) = 2n-2$. Moreover assume there is only one vertex of degree one. Let that vertex be vertex 1 (or rename vertices otherwise). Then $d(1) = 1$ and $d(i) \geq 2$. Then $\sum_i d(i) = d(1) + d(2) + \ldots + d(n) \geq 1 + 2(n-1) = 2n-1$. The latter however is larger than the sum of the degrees $2n-2$, an impossibility.

■

Exercise 533.

The number of binary trees with $n = 1$ is 1. The number of binary trees with $n = 2$ is 2. The number of binary trees with $n = 3$ is 5. Show that the number of binary trees with $n = 4$ is 14. Show that the number is $C(n) = (1/(n+1))\binom{2n}{n}$, the Catalan number of order n.



Solution.
Proven earlier. ■

Exercise 534.
A forest is a graph composed of the union of disjoint trees. Show that an n vertex forest of p trees has $n-p$ edges.

Solution.
Let the vertex disjoint trees be $T_1,\ldots,T_p$, with sizes (number of vertices of each tree) n_i $i=1,\ldots p$ each, such that $\sum_i n_i = n$ (because of the vertexdisjointness). Then, the number of edges in the forest of trees is equal to $\sum_i (n_i-1) = n-p$. ■

Exercise 535.
A graph G that is connected has a spanning tree.

Solution.
If G is a tree (acyclic) then we are done. If G is not a tree, then it must have a cycle. Remove an edge of the cycle. Then remaining graph is still connected. Either it is a tree and has thus a spanning tree or it has a cycle and we repeat the same process. Eventually we reach a graph that is connected and acyclic i.e. a tree. Since it contains all the vertices of (the original G) it is also a spanning tree. ■

Exercise 536.
Show that if G is an undirected connected graph with n vertices and m edges, then G has at least $m-n+1$ cycles. (Two cycles are considered distinct if they differ in at least one edge).

Solution.
Fix a spanning tree of graph G (such a spanning tree exists since G is connected). Let T be this tree. It has $n-1$ edges, and there are $m-n+1$ edges of G not in T. If we add any edge e among these $m-n+1$ edges to T it introduces a cycle (and for different e we get different cycles). Therefore since we have $m-n+1$ different e edges and each one introduces a different cycle, G has at least $m-n+1$ different cycles. ■

Exercise 537.

A connected (undirected) graph is bipartite if an only if it contains no cycles of odd length.

Solution.

If a graph has an odd length cycle it cannot be bipartite: try to color the vertices of the cycle with two colors, red and black where all red vertices are in one bipartition, and black in the other. We are going to bipartition the vertices into two sets R and B.

Pick a vertex u. Start working out a breadth first search from u. All vertices of even length to u (including u) are in R. Those of odd length are in B. No edge $e=(a,b)$ can have both end-points in R or both in B. Say both a,b are in R.

Then $dist(u,a)=2i$ for some integer i, and $dist(u,b)=2j$ for some integer j. But $u\to a\to b$ has $dist(u,b)=2i+1$. Since $2j$ is the shortest length from u to b then $2j\leq 2i+1$. Likewise $2i\leq 2j+1$. Thus $i=j$ since i,j are integer numbers.

Construct the path $p(u,a)$ and $p(u,b)$. Let those two paths be common up to c and deviate from that point. Thus $p(u,c)$ is common to both but $p(u,a)$ and $p(u,b)$ are of the same length and disjoint (if there is an intersection repeat as before and replace c with the new intersection d). Without interesection $p(c,a)$ and $p(c,b)$ along with (a,b) defines an odd length cycle that does not exist. Likewise for the B case. ■

Exercise 538.

Fibonacci trees. Let F_n be the Fibonacci sequence $n \geq 0$, where $F_0 = 0, F_1 = 1$ and thus $F_2 = 1$ as well.

The Fibonacci tree of order n is denoted by T_n or $T(n)$. T_n is a binary tree consisting of a root node, with a left subtree that is a Fibonacci tree of order $n-1$ i.e. T_{n-1} and a right subtree that is a Fibonacci tree of order $n-2$ i.e. T_{n-2}, for $n \geq 3$.
What is the number of nodes $t(n)$ of T_n (i.e. of $T(n)$)?

Solution.
The Fibonacci tree $T(0)$ has 0 nodes. The Fibonacci tree $T(1)$ has 1 node. The Fibonacci tree $T(2)$ has 2 nodes. The Fibinacci tree $T(3)$ has $T(2)+T(1)+1 = 4$ nodes. The Fibinacci tree $T(4)$ has $T(3)+T(2)+1 = 7$ nodes. In general $t(n) = t(n-1)+t(n-2)+1$, if we denote by $t(n)$ the number of nodes for the fibonacci tree $T(n)$. By adding one on both sides we end up with

$$t(n)+1 = t(n-)+1+t(n-2)+1 \Rightarrow P(n) = P(n-1)+P(n-2).$$

The latter is a Fibonacci recurrence. The Fibonacci sequence is $F(n) : 0,1,1,2,3,5,8,\ldots$ $P(n)$ is given however by $t(n)+1 = P(n) : 1,2,3,5,8,\ldots$. Thus $t(n)+1 = P(n) = F(n+2)$ and solving for $t(n)$ we have

$$t(n) = F(n+2)-1.$$

■

Exercise 539.

The following are equivalent for a graph G.
(a) G is 2-colorable, i.e. every vertex is assigned one of two colors so that no edge contains vertices of the same color,
(b) G is biparite, and
(c) every cycle in G has even length (we count edges).

Solution.
Let G be two colorable. Let R and B be the two colors. Color the vertices. Then the R vertices form V_1 and the B vertices form V_2. The induced graph is a bipartite graph. Likewise for a bipartite graph we can color its vertices with two colors.
Let G be a bipartite graph. Let V_1 and V_2 be a bipartition. A cycle that starts from $u \in V_1$ can only then go to a v in $v \in V_2$. If it went to a vertex v such that $v \in V_1$ we would have a violation of the bipartition property. The last edge of the cycle must end to the starting vertex u of V_1. This means the cycle is of even length.
Let every cycle of G is of even length. We pick an arbitrary vertex and we color R. All its neighbors are colored B. If a vertex is B all of its neighbors would be colored R. There is no danger that two adjacent vertices would be colored the same color because that would imply an odd-length cycle. ■

Exercise 540.

A path is an alternating sequence of vertices and edges such that no vertex appears twice. (Implication: no edge appears twice either.)
Given two vertices $u, v \in V$ of a graph $G = (V, E)$, consider all possible paths from u to v and let $l(u, v)$ be the length of the longest one. Define $l(G) = \max_{u,v \in V} l(u, v)$. Show that in a connected graph $G = (V, E)$, any two paths of length $l(G)$ have a vertex in common.

Solution.
Suppose $P_1 = \{u_1, \ldots, u_l\}$ and $P_2 = \{v_1, \ldots, v_l\}$ are two maximum length paths in G that are vertex disjoint. Any pair of vertices in G have a path between them, so we can find a u_i and v_j such that the path between them has no intermediate vertices from either P_1 or P_2.
Assume without loss of generality that $i, j \geq l/2$ (we can always reverse the labelings along the paths to achieve this). Then the walk $u_1, u_2, \ldots, u_i, \ldots, v_j, v_{j-1}, \ldots, v_1$ is a path and has length at least $l + 1$, contradicting our assumption the P_1 and P_2 were maximum. ■

8.4 Eulerican cycles, paths and walks

Exercise 541.
The bridges of Köningsberg (Kalinigrad). Define an undirected graph $G = (V, E)$ with multiple edges such that

$$V = \{1, 2, 3, 4\}$$

and

$$E = \{e_1 = (1,2), e_2 = (1,2), e_3 = (2,3), e_4 = (2,3), e_5 = (1,4), e_6 = (2,4), e_7 = (3,4)\}.$$

The question that the people of Köningsberg asked the famous mathematician Leonhard Euler was the following one.
"Is there a closed walk (tour) of this graph?"
The graph represents the two shores of the city of Köningsberg indicated as vertices 1 and 3, separated by river Pregel and the two islands of Kneiphof and Lopse represented by vertices 2 and 4. In this discussion we assume that the graphs have NO self-loops even if they have multiple edges.

Solution.
Left as a practice problem at the end of this discussion! We show that there is neither an Euler(ian) tour nor an Euler(ian) walk. ■

Exercise 542.

Euler's theorem Version 1. Let $G = (V,E)$ be an undirected graph with multiple edges but no self loops that is connected. Graph G has an Euler tour if and only if every vertex of G is of even degree.
The term Euler tour identifies a closed walk, where every edge is included only once in the walk and vertices may appear more than once.
Note. The problem can be redefined for directed graphs as well with the following modification. For every vertex u the out-degree of u is equal to the in-degree of u. This still leads to the degree of every vertex (sum of in-degree and out-degree) being even.

Solution.
Let G be an undirected graph with multiple edges but no self loops that is connected. We shall show that if G has an Euler tour THEN every vertex of G is of even degree.

If there is an Euler tour T every visit of a vertex v in T that progresses to another vertex includes an in-visit to v followed by an out-visit from v thus consuming two edges out of the degree $d(v)$ of v. Thus if G has an Euler tour the graph must be connected to be able to in-visit and out-visit a given vertex v for all v, and the edges incident on any vertex would be exhausted leading to any v having an even degree. If G were not connected a vertex u would have been unreachable from this tour T and thus there would have been no tour T, contradicting to the existence of the Euler tour T. Moreover, an isolated vertex v would have $d(v) = 0$, which is an even number (so the conclusion would still be true, even if this is a no case since G is connected).
Now, let G be an undirected graph with multiple edges but no self loops that is connected. We shall show that if every verteox of G is of even degree, THEN G has an Euler tour.

We form a walk starting from an arbitrary vertex u. The walk W that is going to be formed would be maximum i.e. we stop when we can have no progress. Let the walk W be then $u, u_1, u_2, \ldots, u_n, v$ and it stops at v where $v \neq u$. We show that this cannot happen. Thus it MUST only stop for $v = u$.
Every vertex u_i in the path may appear once or more. Each appearance leads to an in-visit to that vertex and an out-visit from that vertex: thus two edges of its degree are consumed (note that for this accounting to hold we dismiss self-loops that are not allowed in G though we can incorporate them.) This argument also covers the case of u or v appearing once or more among the u_i's. If the walk stops eventually in v there are two possibilities. (a) v is u itself, and thus edges (u,u_1) and $(u_n,v) = (u_n,u)$ subtract two from the degree of u, and given that we stopped at $v = u$ it means we have used up all edges incident on u. (b) If v is not equal to u, the walk W has only utilized an odd number of edges incident on v (two edges for each occurrence of v as a u_i and one edge (u_n,v) when W stopped at v). But this would imply, given that v is off even degree, that the walk can still progress away from v, and thus the assertion that a maximal W stops at $v \neq u$ should be false: progress can be made to extend W away form u by one more edge thus contradicting to the maximality of W.
Thus W will eventually reach u itself with all edges incident on u exhausted.
There are two possibilities then: either all of the edges of G have been exhausted and W becomes a tour T, or there are some edges not utilized in W. Remove the edges of W from G that is form subgraph G_1 with vertex set V and edge set $E(G) - E(T)$. Every vertex of G_1 has even degree as $E(T)$ contains two edges for every vertex appearance (once or more) in the closed walk W. Note that isolated vertices such as u in G_1 also have even degree (which is zero). Pick a vertex p with degree in G_1 greather than zero and repeat. This might lead to a T_1 that includes all of the remaining edges or not. In the latter case repeat this procedure once more, as needed. In the former case hang T_1 (or connect it) with T. This is possible since G is connected, thus there is a path in G and also in T from u to p, and you attach T_1 to the first instance of the appearance of p in T. Thus any resultant T_i tours (closed walks) can be attached to T. This completes the proof. (We might formulate this proof as an inductive one.) ■

Exercise 543.

This problem is a slight variation of the previous one.

Euler's theorem Version 2. Let $G = (V, E)$ be an undirected graph with multiple edges but no self loops. If graph G has an Euler tour (closed walk where each edge appears once) then G is connected and every vertex of G is of even degree.

Solution.

We first show that if G has an Euler tour, then G is connected.

If there is an Euler tour we isolate for any u and any v the walk from u to v. If it is also a path it means there is a path from u to v for the chosen u and the chosen v. If it is not a path, then it is a walk and it means that there is a (at least one) vertex w that appears multiple times. We remove the edges of the walk between after the first appearance of w through the last appearance of w. (For example: $(1,2,3,4,5,6,3,7,8)$ with $w = 3$ would be reduced to $(1,2,3,7,8)$ by eliminating the cycle $(3,4,5,6)$ i.e. $(3,4,5,6,3)$.) In the remainder w appears only once (the first encounter). The reduced walk (or path) still allows us to go from u to v as one or more cycles or tours were removed. We do so for other vertices q that appear multiple times until we are left with a path from u to v. We repeat this for every u and v. This concludes that G is connected.

We then show that if G has an Euler tour, then every vertex of G is of even degree. (We may also utilize the previous part that showed that G is connected.)

If there is an Euler tour T every visit of a vertex v in T that progresses to another vertex includes an in-visit to v followed by an out-visit from v thus consuming two edges out of the degree $d(v)$ of v. Thus if G has an Euler tour the graph must be connected to be able to in-visit and out-visit a all vertices v of G; moreover the edges incident on any vertex would be exhausted leading to any v having an even degree. If G were not connected a vertex u would have been unreachable from this tour T and thus there would have been no tour T, contradicting to the existence of the Euler tour T. Moreover, an isolated vertex v would have $d(v) = 0$, which is an even number (so the conclusion would still be true, even if this is a no case since G is connected or has been proven to be connected by the first part). If G is connected then there can be no isolated vertex v in G and there can be no vertex v with odd degree. Every visit to an arbitrary vertex v and progress to another vertex away from v consumes two from the degree of v. If there was a vertex v of odd degree, let that v had $d(v) = 2k+1$, an odd number, we could in-visit and out-visit v k times without problems but then the $k+1$-st time an in-visit would lead to getting stuck at v as there would be no unvisited edge for the out-visit, and thus no tour could have been completed. A symmetric case would be a sequence of out-visits followed by an in-visit. The conclusion is that every vertex of even degree greater than zero. This completes the proof. ■

Exercise 544.

Euler's theorem Version 3. An undirected connected graph $G=(V,E)$, with $|V|=n$ and $|E|=m$, has an Euler tour (respectively trail) if and only if none (respectively, exactly two) of its vertices have odd degree.
This implies that all vertices have even degree for an Euler tour or two vertices have odd degree and every other vertex has even degree for a trail and the two vertices of odd-degree are the two end-points of the trail.
In an Euler tour vertices might appear more than once but each and every edge appears exactly one. An Euler trail is an Euler walk where each and every edge appears exactly one and is not closed (i.e. it is not a tour).
Note. The result can be extended for a directed graph. As noted in the previous problem if every vertex has indegree equal to its outdegree then a tour can be found (for a connected graph) or a walk exists if for all but two of the vertices every such vertex has indegree equal to its outdegree, but for the two vertices one has outdegree one more than indegree and the toher indegree one more than outdegree.

Solution.
1a. Necessary part: Euler tour case. If there is an Euler tour, then this means that the graph is connected by the Euler's theorem version 2 problem shown previously. Moreover we showed in the previous problem that every vertex has even degree.
1b. Necessary part: Euler trail case. If there is an Euler trail, then this means that the graph is connected. It can be shown in the same way the previous result (Euler's theorem version 2 problem) was shown. Moreover we showed in the previous result that every vertex has even degree for the case of an Euler tour. But in case of a trail the starting and ending vertices must be of an odd degree.
Therefore in the case of an Euler tour or Euler trail in a connected undirected graph either all vertices have even degree or two vertices have odd degree and the remaining ones have even degree. The two vertices with odd degree are the starting and ending vertices of the Euler walk.
2. Sufficient part.
The case that all vertices are of even degree has been proven in the Euler's theorem version 2 problem and thus could have been omitted. We show by induction the following.

If there are only two vertices u and v with odd degree then there exists an Euler trail that starts from u and ends in v (or the other way around); if there are no vertices with odd degree, then the graph has an Euler tour.

We prove this by induction on m, the number of edges of G.

We shall assume that the statement above is true for all graphs with k edges, where $k<m$, and show that it is also true for a graph G with m edges (strong induction).
We assume that there are two vertices u,v of odd degree. The argument reminds us of the second part of the Euler's theorem version 1 proof.
Let us start with u to create a walk W. We move along by never visiting an edge twice. If we reach $q\neq v$, through an edge there is a way out of it through another edge since q has even degree. If $d(q)=2k$ and we had visited q $k-1$ times there are two unvisited edges, one on the way into q and one on the way out of q. Note that q can include u and v as well. When it is no longer possible to move, it means we have reached v. (Any other node has even degree thus a way to reach that node, should allow for an exit from that node as well.) It is possible then that not all edges have been visited (used) through walk W.
After we remove the visited (used) edges from G, the remainder of the graph is such that every vertex has even degree. Let $G_1,G_2,\ldots,G_k$ be the connected components of G that have at least one edge (i.e. at least two vertices). By induction they have Euler tours $t_1,t_2,\ldots,t_k$. Since G is connected the original walk W encounters all of $G_1,\ldots,G_k$. Let W encounters G_1 at vertex r_1, G_2 at vertex r_2, and so on. Then we generate the walk by combining W, $t_1,\ldots,t_k$.

$$W(u\rightsquigarrow x_1)\rightsquigarrow t_1\rightsquigarrow W(x_1\rightsquigarrow x_2)\rightsquigarrow t_2\rightsquigarrow \ldots W(x_k\rightsquigarrow v)$$

The resulting walk is an Euler walk. We can work for an Euler tour similarly of rely to the second part of the proof of Euler's theorem version 1. The resulting tour is an Euler tour then. ■

Exercise 545.
If G is a connected undirected graph (with or without multiple edges) and every vertex has even degree, then G has an Euler tour.

Solution.
Direct consequence of (a) previous theorem(s). ■

Exercise 546.

Euler's theorem on directed Graphs. A connected directed graph G has an Euler tour if and only if each vertex of it has in-degree equal to its out-degree.

Solution.
Implied by a prior discussion and notes. ■

Exercise 547.

Show that if G is a connected undirected graph with k vertices of odd degree $(k > 0)$ then there are $k/2$ walks no two of which share an edge, that between them contain all the edges of G.

Solution.
First, one must show that the problem is well defined in the sense that k is always an even number. This follows form the following claim.

The number of odd-degree vertices in an undirected graph $G = (V, E)$ is even.
This can be proven by a simple counting argument relying on the fact that the sum of the degrees of the vertices is two times the number of the edges and thus an even number.

Therefore, we proceed based on k being an even number. Let $k = 2m$.
We divide the k vertices into m pairs of vertices and we form a new graph G' by adding to G m new edges, each joining a pair of the odd degree vertices (even if such edges already exist). Then we have that G' is a connected graph, all of whose vertices have even degree. From Euler's theorem, we get that there's a tour T in G' that traverses each edge exactly once. We now remove from T the m edges we added, breaking the tour into m segments such that the remaining edges generate m trails that use all edges of G each trail starting from an odd-degree vertex and ending in another odd-degree vertex. In any collection of trails using all edges of G, every odd vertex must be the end-point of a walk and so with k odd-degree vertices we cannot cover all edges of G using less than m paths. ■

Exercise 548.
The bridges of Königsberg (Kalinigrad) revisited.

Solution.
To conclude this problem the bridges of Königsberg graph is a graph where all 4 vertices are of odd degree. Thus there is neither an Euler trail nor an Euler tour. Concluded. ■

Exercise 549.

De Bruijn sequence. A De Bruijn sequence of order k from an alphabet $\Sigma = \{0,1\}$ of cardinality two, is a cycle B_n (also indicated $B(k)$) of $n = 2^k$ zeroes and ones that generate all binary strings of length k.
Give an De Bruijn sequence of order 3.
Give an De Bruijn sequence of order 4.
Is 0000111101100101 a De Bruijn sequence of order 4?

Solution.
(a) A De Bruijn sequence of order 3 is of length $2^3 = 8$ and is shown below. We write it down flatly yet because it is a cycle a wrap-around is implied from it last bit which is one to its first bit which is a zero.

$$00011101$$

It is quite clear that in a cycle form (clockwise) it can be written as follows. Sequence B_k is usually written in normalized form to start with a 000.

```
      0          (a)  B(n) for n=3
    1   0
   0     0
    1   1
      1

 (0 0 0)1 1 1 0 1   -->  000
  0(0 0 1)1 1 0 1   -->  001
  0 0(0 1 1)1 0 1   -->  011
  0 0 0(1 1 1)0 1   -->  111
  0 0 0 1(1 1 0)1   -->  110
  0 0 0 1 1(1 0 1)  -->  101
  0)0 0 1 1 1(0 1   -->  010
  0 0)0 1 1 1 0(1   -->  100
```

Note that B_1 is 01, B_2 is 0011. Moreover besides 00011101 there is also 00010111. An of course the reverse of 00011101 which is 10111000 is also a De Bruijm sequence. One can show that the number of $B(n)$ sequences is equal to $2^{2^{n-1}-}$. The result was proven by N. De Bruijn.
(b) An example of a B_4 is given below.

$$0000110100101111$$

(c) Omitted. Is 0000110100101111 a De Bruijn sequence? ■

Exercise 550.

De Bruijn graph of order k. A De Bruijn graph $D(k)$ of order $k=3$ is defined as a directed graph $G=(V,E)$ where the vertices are k-bit sequences. Thus $|V|=n=2^k$. There is an edge from $a_1a_2a_3$ to $a_2a_3a_4$, for all $0\leq a_i\leq 1$. In particular from $a_1a_2a_3$ there is a zero edge leading to a_2a_30 and an one edge leading to a_2a_31. Thus for the $k=3$ case the vertex set is

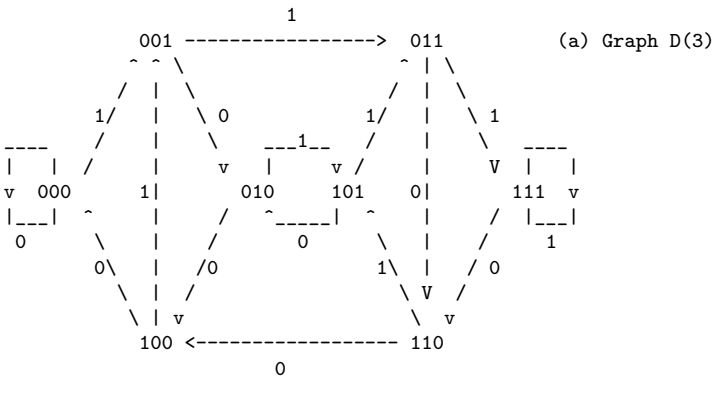
$$V=\{000,001,010,011,100,101,110,111\}.$$

Moreover, the edge set E is

$$\begin{aligned} E &= \{(000,000),(000,001),(001,010),(001,011),(010,100),(010,101),(011,110),(011,111),\\ &\quad (100,000),(100,001),(101,010),(101,011),(110,100),(110,101),(111,110),(111,111),\}. \end{aligned}$$

The edge $(000,000)$ can be defined as $000-0$, where as edge $(010,101)$ as $010-1$.
(a) Show that the graph has an Euler tour.
(b) Show that the graph has a Hamiltonian cycle.
(c) Conclude that a De Bruijn sequence of order k can be obtained as either a Hamiltonian cycle or a Euler tour of a De Bruijn graph of some order, whether it is k or $k-1$.

Solution.
(a) The De Bruijn graph $D(3)$ is show below. Every vertex has in-degree 2 and an out-degree 2. Therefore it has an Euler tour. The self loop is considered an edge that contributes $+1$ to the out-degree and $+1$ to the in-degree of vertex 000 and 111. Moreover an Euler tour with out the two-self loops can be extended into an Euler tour with the two self loops $(000,000)$ and $(111,111)$ quite obviously.



Note that a vertex of $D(3)$ maps to an edge of $D(2)$. Thus edge $00-0$ is vertex 000 of $D(3)$ and edge $00-1$ of $D(2)$ becomes vertex 001 of $D(3)$. Consider the De Bruijn sequence B_4 from the previous problem.

$$0000110100101111$$

It describes an Euler tour on $D(3)$ if every four bit strings abcd (in circular order) are interpreted as the edge abc-d.

```
                                      abc    bcd  d
(0 0 0 0)1 1 0 1 0 0 1 0 1 1 1 1   000 -> 000  0
 0(0 0 0 1)1 0 1 0 0 1 0 1 1 1 1   000 -> 001  1
 0 0(0 0 1 1)0 1 0 0 1 0 1 1 1 1   001 -> 011  1
 0 0 0(0 1 1 0)1 0 0 1 0 1 1 1 1   011 -> 110  0
 0 0 0 0(1 1 0 1)0 0 1 0 1 1 1 1   110 -> 101  1
 0 0 0 0 1(1 0 1 0)0 1 0 1 1 1 1   101 -> 010  0
 0 0 0 0 1 1(0 1 0 0)1 0 1 1 1 1   010 -> 100  0
 0 0 0 0 1 1 0(1 0 0 1)0 1 1 1 1   100 -> 001  1
 0 0 0 0 1 1 0 1(0 0 1 0)1 1 1 1   001 -> 010  0
 0 0 0 0 1 1 0 1 0(0 1 0 1)1 1 1   010 -> 101  1
 0 0 0 0 1 1 0 1 0 0(1 0 1 1)1 1   101 -> 011  1
 0 0 0 0 1 1 0 1 0 0 1(0 1 1 1)1   011 -> 111  1
 0 0 0 0 1 1 0 1 0 0 1 0(1 1 1 1)  111 -> 111  1
 0)0 0 0 1 1 0 1 0 0 1 0 1(1 1 1   111 -> 110  0
 0 0)0 0 1 1 0 1 0 0 1 0 1 1(1 1   110 -> 100  0
 0 0 0)0 1 1 0 1 0 0 1 0 1 1 1(1   100 -> 000  0
```

Moreover the d values of the edges form a De Bruijn sequence. (b) The De Bruijn sequence $B(3) = 00011101$ leads into a Hamiltonian cycle.

```
(0 0 0)1 1 1 0 1   -->  000
 0(0 0 1)1 1 0 1   -->  001
 0 0(0 1 1)1 0 1   -->  011
 0 0 0(1 1 1)0 1   -->  111
 0 0 0 1(1 1 0)1   -->  110
 0 0 0 1 1(1 0 1)  -->  101
 0)0 0 1 1 1(0 1   -->  010
 0 0)0 1 1 1 0(1   -->  100
```

The list of eight vertices define a Hamiltonian cycle's first seven edges with the last eight edge being $(100,000)$ implied between and the last and the first vertex respectively.
On B_3 consider reading the labels of the edges when you visit vertices once: 10111000 or 01000111 or 11101000.
■

Exercise 551.

Generalized De Bruijn graph $D(h,k)$. Let $h,\ k \geq 2$ be natural numbers and set $V = \{0,1,\ldots,h-1\}^k$. The generalized De Bruijn graph $D(h,k)$ has vertex set V and every vertex a $= (a_1,a_2,\ldots,a_k)$ is adjacent to every vertex b $= (a_2,a_3,\ldots,a_k,*)$, where $*$ denotes an arbitrary element of $\{0,1,\ldots h-1\}$. The following is immediate from the definition.

Fact 8.1

The Generalized De Bruijn graph $D(h,k)$ has order i.e. number of vertices equal to h^k and for all $u \in V(D(h,k))$, indegree (u) = outdegree (u) = h.

Since for every vertex $indegree = outdegree$ we conclude that there exists an Euler tour in $B(h,k)$.
Using an Euler tour argument to find a circular arrangement of nine A's, nine B's and nine C's such that each of the 27 subsequences of length 3 are distinct. Show that arrangement!

Solution.
For the particular solution of this problem, we need to construct $D(3,2)$ where $V = \{A,B,C\}^2$ and find an Euler tour in that graph.
The following shows a circular arrangement with the desired properties (with 9 A's, 9 B 's , 9 C's):
AAACACBCCBBBCBAABABBACCCABC ■

Exercise 552.
How many different (up to rotation) De Bruijn sequences of length 8 are there? Construct a De Bruijn sequence of length 32 (Use Euler walk arguments in each case).

Solution.
The number of distinct De Bruijn sequences of length 8 (excluding those obtainable by rotations) corresponds to the number of closed Euler walks in the four vertex De Bruijn graph $D(2)$. This is seen to be two (11101000 and 11100010), depending on whether the loop between the 01 and 10 vertices is taken before or after visiting the 11 vertex. ■

Exercise 553.
On a 3×3 chessboard there are 2 white horses (knights) stationed at the bottom corners and two black horses (knights) stationed at the top corners. Horses are moved in an L shape of two up or two down moves and one left or right move, or two left or two right followed by an one up or one down move. Is there a way of legal chess movements that would allow the horses to swap possitions? (That is the black horses move to the bottom and white horses to the top corners?)

Solution.
Label the 9 positions $1 \ldots 9$ row-major or column-major or snake-like. These are the vertices of a graph. Join i and j with an edge if there is a (as in one) horse L movement from i to j and for an undirected graph that it will be from j to i. The middle square is unused and can be thrown away. The rest (the subgraph excluding the middle square) has the property that every vertex is of even degree. Thus there is an Euler tour. (For example move the horses on a cyclical pattern of the squares.) Further details omitted. ■

Exercise 554.
Show that all De Bruijn graphs have Hamiltonian cycles. A De Bruijn graph is defined as $G = \{V, E\}$ where $V = \{\alpha_1 \dots \alpha_n \mid \alpha_i \in \{0,1\}\}$ and $E = \{(\alpha_1 \dots \alpha_n, \alpha_2 \dots \alpha_{n+1}) \mid \alpha_i \in \{0,1\}\}$.

Solution.
Let G_n be the De Bruijn graph of 2^n vertices where an Euler cycle corresponds to a De Bruijn sequence of all strings in $\{0,1\}^{n+1}$. In the special case where $n = 1$ this graph clearly has a Hamiltonian cycle $(0 \to 1 \to 0)$. Otherwise, consider a consider the De Bruijn sequence that includes all strings in $\{0,1\}^n$. Each substring of n bits from that sequence is a vertexin G_n. Every two adjacent substrings $\alpha_1 \dots \alpha_n$ and $\alpha_2 \dots \alpha_{n+1}$ are vertices that are adjacent in n. Thus this De Bruijn sequence describes a sequence of edges in G_n that is a walk, touches every vertex(since every substring is in the sequence), touches each vertexexactly once (no substring appears more than once), and is closed. This is a Hamiltonian cycle. ■

Exercise 555.
Show that for every undirected multigraph $G = (V, E)$ there exists a closed walk that covers every edge of the graph exactly twice.

Solution.
Take G and create a modified multigraph $G' = (V, E')$ where E' is just E with every edge repeated. Clearly every vertex in G' has even degree. G' therefore has a closed Euler walk, which when mapped back to G crosses every edge exactly twice. ■

Exercise 556.
Let G be a graph with every vertex of even degree. Then graph G is an edge-disjoint union of cycles.

Solution.
Induction-based on m, which is the number of edges. True for $m = 2$ (undirected has multiple edges; directed case does not). Suppose this is true for all $m < k$. Pick a graph with $m = k$ edges. Start with u. There is an outgoing edge leading to v, then visit v. Since degree of v is two there is another edge that is followed as an outgoing edge from v. Continue until you end up back to a vertex already visited call it w. Then the path from w to w again is a cycle. Remove it from the graph (every vertex of the cycle has 2 removed from its degree that remains even). The graph has $< k$ edges. Use induction. (No need to determine if there are multiple connected components and induct on them as well.) ■

8.5 Hamiltonian cycles and paths

Exercise 557.

In an undirected graph $G = (V, E)$ that is regular of degree 3, the number of Hamiltonian cycles that contain a given edge is an even number.

Solution.

Since every vertex has degree 3, an odd number, the number of vertices is even so that the sum of the degrees is still an even number and twice the number of edges. Since n is even the Hamiltonian cycle is of even length and thus induces a coloring of the edges of the cycle as B and W; the remaining non-cycle edges are colored G. Every vertex has degree 3; thus the set of edges of each one of three colors generate a perfect matching. ■

Exercise 558.

Auxiliary Hamiltonian Cycle Result (AHCR).
Let G be an undirected graph $G=(V,E)$. Let $d(u)+d(v)\geq n$ for two non-adjacent vertices. If $G+(u,v)$ is Hamiltonian (has a hamiltonian cycle), then G is also Hamiltonian.
$G+(u,v)$ denotes the graph G with updated E to include edge (u,v) that is, $E=E\cup\{(u,v)\}$.

Solution.
Suppose that $G+(u,v)$ has a Hamiltonian cycle c. If c does not utilize edge (u,v) then this means c is also a Hamiltonian cycle in G and we are done.
If c utilizes edge u,v, then c contains a path p such that $p=c-(u,v)$. For the sake of simplicity let us relabel the vertices of the cycle/path as follows is $u_1=u,u_2,u_3,\ldots,u_{n-2},u_{n-1},u_n=v$.
We form two sets

$$F=\{u_i:(u,u_{i-1})\in E,\ \textbf{and}3\leq i\leq n-1\}$$

$$G=\{u_i:(v,u_i)\in E,\ \textbf{and}3\leq i\leq n-1\}$$

utilizing

$$H=\{u_3,\ldots,u_{n-1}\}$$

which contains $n-3$ vertices. Since $d(u)+d(v)\geq n$ we have

$$|F|+|G|=d(u)-1+d(v)-1=d(u)+d(v)-2>n-2.$$

by the pigeonhole principle there must be an i i.e. $u_i\in F$ and $u_i\in G$. Then we can do the following to generate a new Hamiltonian cycle

$$u_i\rightsquigarrow u_1=u\rightsquigarrow u_2\rightsquigarrow u_3\ldots\rightsquigarrow u_{i-1}\rightsquigarrow u_n=v\rightsquigarrow u_{n-1}\rightsquigarrow\ldots\rightsquigarrow u_i$$

by reusing segments of the

$$u_1=u\rightsquigarrow u,u_2\rightsquigarrow u_3\ldots\rightsquigarrow u_{n-2}\rightsquigarrow u_{n-1}\rightsquigarrow u_n=v.$$

A Hamiltonian cycle has been obtained for G from the Hamiltonian cycle c of $G+(u,v)$. ■

Exercise 559.

Dirac's Theorem. In an undirected graph $G = (V, E)$, with $n \geq 3$, if every vertex u has $d(u) \geq n/2$, then G has a Hamiltonian cycle.

Solution.
Note that for every two vertices a and b since $d(a) \geq n/2$ and $d(b) \geq n/2$ for all $a, b \in V$ we also have $d(a) + d(b) \geq n$.
Case 1. G has a Hamiltonian cycle. Done.
Case 2. G does not have a Hamiltonian cycle. Utilizing AHCR, if there are two vertices u, v unconnected, we explore whether $G + (u, v)$ has a Hamiltonian cycle. If it does by AHCR so should G contradicting the "G does not have a Hamiltonian cycle". Thus we go on adding edges to unconnected vertices in G until it eventually gets a Hamiltonian cycle (in the worst case it becomes K_n). Then we undo the addition of edges using AHCR to show that G must have a Hamiltonian cycle contradicting the assumption that "G does not have a Hamiltonian cycle". Thus Case 2 does not exist. ■

Exercise 560.

Prove the following version of Dirac's theorem. If we have an undirected connected graph $G = (V, E)$ such that for any two non-adjacent vertices u, v we have that $d(u) + d(v) \geq n$ ($d(w)$ is the degree of a vertex w in G), then G must have a Hamiltonian cycle.

Solution.

The argument is virtually identical to that used for Dirac's theorem. Add additional edges to G until no more can be added without forming a Hamiltonian cycle. Let x and y be two non-adjacent vertices. There must be a Hamiltonian path $x = z_1 \to z_2 \to \ldots \to z_{n-1} \to z_n = y$. If for some $2 \leq j \leq n$ we have x adjacent to z_j and y adjacent to z_{j-1} then $x = z_1 \to z_2 \to \ldots \to z_{j-1} \to z_n = y \to z_{n-1} \to \ldots \to z_j \to z_1 = x$ is a Hamiltonian cycle. But if there is no such j then the degree of y can be at most $(n-1) - d(x)$ which contradicts the assumption that $d(x) + d(y) \geq n$.

■

Exercise 561.
Show that if every vertex in an n vertex graph (where n is even) has degree at least $n/2$, then the graph has a perfect matching.

Solution.
By Dirac's theorem, such a graph has a Hamiltonian cycle of length n. Taking every second edge from this cycle gives a matching, since no vertex will be included more than once. Because n is even, every vertex will be included once, so the matching is perfect. ■

Exercise 562.

Let $G = (V, E)$ be an undirected connected graph such that for every two vertices u, v we have that $d(u) + d(v) \geq n$ ($| V | = n$), where $d(w)$ is the degree of vertex w. Show that for an even n, G has a perfect matching (i.e. a complete matching that saturates all the vertices of the graph).

Solution.

One must first prove that under the restriction that $d(u) + d(v) \geq n$, graph G must have a Hamiltonian cycle. The proof is identical to the Dirac's Theorem proof (note that there, we only needed the fact that the sum of the degrees of two non-adjacent vertices is at least n, and this condition holds for this problem too). Therefore, G has a Hamiltonian cycle and since it has an even number of vertices, we can find a perfect matching by deleting from the cycle every other edge. What is left is a matching that saturates all the vertices of G, i.e. a perfect matching (and what has been deleted is a second perfect matching!). ■

Exercise 563.

Show that if $G=(V,E)$ is an undirected graph with $|V|=n$ and $|E|=m$ and $m \geq \frac{1}{2}(n-1)(n-2)+2$, then G must have a Hamiltonian cycle.

Solution.

We use in this problem Dirac's theorem and in particular that if for all pairs of non-adjacent vertices u,v of G we have $d(u)+d(v) \geq n$, then G must be a Hamiltonian graph (i.e. it must have a Hamiltonian cycle).
Suppose we have a graph G which has at least $\frac{1}{2}(n-1)(n-2)+2$ vertices. We prove that way.
Assume for contradiction that the graph is not Hamiltonian. Then there must exist a pair of non-adjacent vertices such that $d(u)+d(v)<n$ since, otherwise all pairs of non-adjacent vertices are as in the statement of a prior problem and therefore a Hamiltonian cycle exists, a contradiction to our assumption. We now find an upper bound on the number of edges G might have. The $n-2$ vertices of G other than u,v may be at most connected in a complete graph fashion and among them we might have at most $\frac{1}{2}(n-2)(n-3)$ edges. Vertex u is connected to $d(u)$ of these vertices and v to $d(v)$, since u and v are not connected each other. Therefore, the maximum number of edges in G is

$$|E| \leq \frac{1}{2}(n-2)(n-3)+d(u)+d(v) \leq \frac{1}{2}(n-2)(n-3)+n-1 \leq \frac{1}{2}(n-1)(n-2)+1$$

since we assumed that $d(u)+d(v)<n \Rightarrow d(u)+d(v) \leq n-1$. But we have at least $\frac{1}{2}(n-1)(n-2)+2$ edges, a contradiction. Therefore no pair of non-adjacent vertices with $d(u)+d(v)<n$ exists, and the graph has thus a hamiltonian cycle. ■

Exercise 564.

A set of $2k+1$ people plan to have dinner together around a circular table on k successive nights in such a way that no pair sit next to each other more than once. Show that this is possible for $k=5$.

Solution.
For the case $k=5$ one can find the following ordering of the people around a circular table (assuming a circular order of the lists below)

- First night: $1,2,3,4,5,6,7,8,9,10,11$.
- Second night: $1,3,5,7,9,11,2,4,6,8,10$.
- Third night: $1,4,7,10,2,5,8,11,3,6,9$.
- Fourth night: $1,5,9,2,6,10,3,7,11,4,8$.
- Fifth night: $1,6,11,5,10,4,9,3,8,2,7$.

We now show how we can find k circular orderings of the $2k+1$ people for any k. Initially, each person can sit next to anyone else. We represent this relationship with a graph. The vertices of the graph are the people, and we have an edge between a and b if a can be seated next to b. The graph constructed is the K_{2k+1} (complete graph on $2k+1$ vertices). Here's is how we can find the k orderings. We put $2k$ of the points on a circle and the other point in the center, and we construct the initial ordering shown there. Figure 1 below illustrates this construction. Figure 2 shows the initial ordering for $k=4$ while Figure 3 shows how we can get a new ordering (edge disjoint) by a right one step rotation of Figure 2. Rotating Figure 1 clockwise, in the sense shown in Figure 3, by $1,2,\ldots k$ steps, we get all possible orderings we need..
We now only need to prove that all these orderings do not use the same edge twice, i.e. that the k hamiltonian cycles of the K_{2k+1} complete graph generated by these rotations are edge-disjoint, and so the traversal of the vertices of each cycle gives a valid ordering. One can see, using the labelling of the vertices in Figure 1, that two vertices (on the circle) are adjacent in Figure 1 if their sum is 2 or 3 modulo $2k$. If we rotate this first ordering clockwise by one step the sum of two adjacent vertices on the circle becomes 4, or 5 modulo $2k$. Another rotation gives a sum of 6, 7 modulo $2k$ and so on. It is easy to see, that the "diameter edges" are disjoint for these k orderings. Therefore the rotation of the original ordering can be repeated a total of $k-1$ times before getting the original one. Hence, in that way we get k edge-disjoint Hamiltonian cycles i.e. the orderings of people we would like to have . ■

Exercise 565.
Show that binary Gray codes exist for every n (n is the length of a word of the code).

Solution.
This is shown by induction on n. The basis where $n = 1$ is trivial as $0, 1$ suffices. Now assume the hypothesis is true for $n = k$. Let $u_1, u_2, \ldots, u_{2^k}$ be a Gray code. Construct the code $0u_1, 0u_2, \ldots, 0u_{2^k}, 1u_{2^k}, 1u_{2^k-1}, \ldots, 1u_1$. This is a Gray code for $n = k+1$ since each pair of adjacent strings differ in only one bit. This is shown by induction on n. The basis where $n = 1$ is trivial as $0, 1$ suffices. Now assume the hypothesis is true for $n = k$. Let $u_1, u_2, \ldots, u_{2^k}$ be a Gray code. Construct the code $0u_1, 0u_2, \ldots, 0u_{2^k}, 1u_{2^k}, 1u_{2^k-1}, \ldots, 1u_1$. This is a Gray code for $n = k+1$ since each pair of adjacent strings differ in only one bit. ■

Exercise 566.
Show that there is an ordering of the decimal numbers of length n $(00\ldots0-99\ldots9)$ where each pair of cyclically adjacent numbers differs in only one digit.

Solution.
This is a slight generalization to the previous problem, which again we show by induction on n. If the hypothesis is true for $n=k$ let $u_1,u_2,\ldots,u_{10^k}$ be a Gray code. Construct the code that begins $0u_1,0u_2,\ldots,0u_{10^k}$, $1u_{10^k},1u_{2\,10^k-1},\ldots,1u_1$, continues $2u_1,2u_2,\ldots,2u_{10^k}$, $3u_{10^k},3u_{2\,10^k-1},\ldots,3u_1$, then has $4u_1,4u_2,\ldots,4u_{10^k}$, $5u_{10^k},5u_{2\,10^k-1},\ldots,5u_1$, and so on up to $8u_1,8u_2,\ldots,8u_{10^k}$, $9u_{10^k},9u_{2\,10^k-1},\ldots,9u_1$, This is a Gray code for $n=k+1$ since each pair of adjacent strings differ in only one bit. ■

Exercise 567.
On an 8×8 chessboard, is it possible for a horse (knight) to visit each square of the chessboard only once AND return to the original (home) square?

Solution.
Check for a Hamiltonian cycle on a graph of horse movements. Hamilton and Vandermonde provided solutions. For $n = 4$ no solution exists and also, for n odd. ■

Exercise 568.

List a permutation of the first n positive integers $n \geq 2$ so that consecutive integers have a sum that is a integer power (integer power of an integer).
For example 81726354 is such since $8+1$, $1+7$, $7+2$, $2+6$, e.t.c. are powers of 3, 2, 3, 2, e.t.c. respectively.
Can you find one for $n = 7$?
Hint: Start with 4.

Solution.
4 is a starting number that has a unique next number: 5. Thus 453 is easily determined and to conclude 4531726 or 4536271 are possibilities. Draw a graph and compute a Hamiltonian path otherwise. ■

Exercise 569.

Given a graph G, the weight of a path is the sum of the weights of edges belonging to that path.
Define the MAXIMUM PATH problem as follows: Given a graph G, find a simple path (i.e. a path with no cycles) that has maximum weight.

(i) Give an efficient algorithm for the MAXIMUM PATH problem on acyclic graphs.
(ii) It is believed that no efficient algorithm exists for testing whether an arbitrary graph has a Hamiltonian path. Based on this fact show that you don't expect there to be an efficient algorithm that solves the MAXIMUM PATH problem in the general case.

Solution.
(i) Apply Floyd's Algorithm to our graph (it doesn't matter if we have negative weights since no problem-causing negative weight cycles exist). Since the graph is acyclic a path between two vertices is maximal and minimal simultaneously, so the shortest path algorithm is a maximum path algorithm as well for acyclic graphs.
(ii) Take an arbitrary graph, put weight 1 on each edge of it, and if a MAXIMUM PATH algorithm for general graphs exists apply this algorithm to this graph. If the graph has a hamiltonian path between two of its vertices then the MAXIMUM PATH algorithm will return a path with weight $n-1$ (if MAXIMUM PATH gives a path of weight $< n-1$ then no Hamiltonian path exists). Therefore, an efficient algorithm for the MAXIMUM PATH problem for general graphs, will give immediately an efficient algorithm for the Hamiltonian problem (but we believe no such efficient algorithm exists). ■

Exercise 570.

Show that every regular, degree 3, Hamiltonian graph has exactly 3 edge-disjoint perfect matchings. A graph is Hamiltonian if it possesses a Hamiltonian cycle.

Solution.

Let $G = (V, E)$ be a 3-regular, Hamiltonian graph. Since every vertex has degree 3 and the sum of the degrees of all the vertices is an even number, the number of vertices of the graph must be even. We know the graph has a Hamiltonian cycle, let it be C. From this Hamiltonian cycle, we can get 2 edge disjoint matchings, one by taking every other edge of the cycle (and we saturate every vertex because we have an even number of them), the other by taking the edges that are left on the cycle. By deleting C from G, we decreased the degree of every vertex by 2, i.e. every vertex has degree 1, and this is the third matching, edge disjoint from any previous one. Since the degree of every vertex is 3, we can have at most 3 edge disjoint perfect matchings. ■

Exercise 571.

A chess King is on the upper left square of an 8x8 chessboard (shown below), and a Queen is on the lower right square. The King wishes to visit the Queen at the lower right square, and he can move to any adjacent square but cannot move diagonally. Is it possible for him to visit each square of the chessboard exactly once, en route to the Queen? Explain your answer in graph theoretic terms.

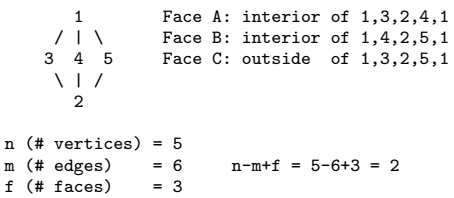
Solution.
We can find a solution to this problem, if there exists a hamiltonian path from the upper left corner to the lower right corner of the chessboard. We get a graph $G = (V, E)$ modelling the problem by representing every square with a vertex and the at most four possible moves from a square with edges to adjacent vertices. From the chessboard, we can see, that a move from one square to its adjacent ones flips the color of the square (diagonal moves are not allowed). If there were to exist a Hamiltonian path from King's initial position to the Queen's position it should have been of length 63. But after 63 moves (an odd number) the color of King's initial square is flipped, i.e. the King must be on a black square (since the color of its initial square was white), but the Queen is sitting on a white square. Therefore it's not possible for the King to meet the Queen and visit all the squares of the chessboard exactly once en route to the Queen (when diagonal moves are not allowed). ■

8.6 Planar graphs

Exercise 572.

For the graph $K_{2,3}$ with $V_1 = \{1,2\}$ and $V_2 = \{3,4,5\}$, show that it is planar give its faces and show that $n-m+f=2$ where f is its number of faces.

Solution.

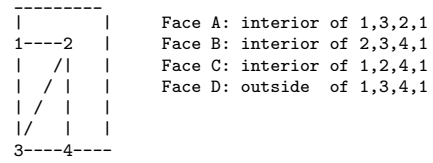



```
n (# vertices) = 5
m (# edges)    = 6     n-m+f = 5-6+3 = 2
f (# faces)    = 3
```

■

Exercise 573.
Likewise for K_4.

Solution.




```
n (# vertices) = 4
m (# edges)    = 6     n-m+f = 4-6+4 = 2
f (# faces)    = 4
```

■

Exercise 574.

(Yet another) Euler's theorem. A planar graph with n vertices, m edges and f faces satisfies the following equality.

$$n-m+f=2.$$

Prove the result by induction.

Solution.
For $n=1$, $m=0$ we have $f=1$ obviously. For $n=2$, $m=1$ we have $f=1$ obviously. Both by inspection.
Induction on m. Assume it is true for all planar graphs with m edges of less.
Pick a planar graph with $m+1$ edges and pick one of the m edges call it e.
Case 1: Edge e connects an isolated vertex to G.
Graph G let us say it has n vertices $m+1$ edges and f faces. We want to calculate $n-(m+1)+f$ and confirm it is equal to two. Remove e i.e. $G-e$. Graph has $n-1$ vertices m edges and the same number of faces as before. Moreover it is still planar. Thus by induction $(n-1)-m+f=2$. Then $n-(m+1)+f=(n-1)-m+f=2$ and the result has been proven.

Case 2. Edge e connects two vertices that are not isolated.
Graph G let us say it has n vertices $m+1$ edges and f faces. We want to calculate $n-(m+1)+f$ and confirm it is equal to two.
Graph $G-u$ in this case still has n vertices m edges. The removal of e joins two faces on either side into one thus the number of faces becomes $f-1$. Therefore by induction

$$n-m+(f-1)=2 \Rightarrow n-(m+1)+f=2,$$

and the result has been proven. Proof is thus complete by case analysis. ■

Exercise 575.

Euler's formula. For a tree T we have the following $n-m+f=2$, where n is the number of vertices, m the number of edges and f the number of faces.

Solution.

We start from a generic planar graph G. If the graph has a cycle remove an edge from it. As a result, the new value of f is the previous value of f minus 1 and the new m is the previous value of m minus 1. So the resulting graph has $N=n, M=m-1$ and $F=f-1$. Thus $N-M+F=n-m+1+f-1=n-m+f$. Repeat until no cycles remain. The resulting graph is a tree and thus $N-M+F=n-(n-1)+1=2$. Moreover, we can prove the same result directly: a tree is an "open" structure without cycles and with n vertices and $n-1$ edges. Because it is open there is only one face in it. ■

Exercise 576.
K_5 is not planar. Prove it.

Solution.
Say K_5 is a planar graph. K_5 has $n=5$ and $m=10$ and then $n-m+f=2$ implies $f=7$.
Each face must be surrounded by at least 3 edges. Let then b be the number of boundarye (edges) of those faces i.e. $b \geq 3f$. Moreover $b=2m$ since an edge is used as a boundary twice (one side is a face, the other side is another face).
Then $b=2m \geq 3f$ implies $m \geq 3/2f$. But then $10 \ngeq 3/2 \cdot 7$, since $m=10$ and $f=7$. ■

Exercise 577.
Prove that $K_{3,3}$ is not planar.

Solution.
Omitted. ■

Exercise 578.
Show that in any connected planar graph with at least two faces the following is true: $m \leq 3n-6$.

Solution.
Every face has at least 3 edges. Each edge is shared by two faces As a result of these considerations, $2m \geq 3f$ and therefore solving for m we have: $m \geq 3f/2$.
Using $n-m+f=2$ we have $f=2-n+m$ and $m \geq 3(2-n+m)/2$ which leads to

$$m \leq 3n-6,$$

as required. ■

Exercise 579.

Reminder. A graph is planar if and only if it contains no homomorphic emdedding of K_5 or $K_{3,3}$.
A planar embedding of a graph is a drawing of a graph in the plane with out crossing edges. A graph $G_1 = (V_1, E_1)$ is homomorphically embeded into graph $G_2 = (V_2, E_2)$ if there is an 1-1 mapping $p : V_1 \to V_2$ such that for all edges $(u, v) \in E_1$, edge (u, v) of G_1 maps to path $p(u) \rightsquigarrow p(v)$ of G_2 such that the paths are vertex disjoints except for the end-points.

Solution.
Omitted. ■

Exercise 580.
(i) Show that the number of faces in a planar graph is at most $2n-4$, where n is the number of vertices.

(ii) Show that if a planar graph has 6 vertices and 12 edges then every face is bounded by exactly 3 edges.

Solution.
(i) Recall Euler's theorem that any planar embedding of a *connected* graph with e edges and n vertices has f faces where

$$f = e - n + 2$$

In a graph with k connected components, then this is true for each component. Summing, we get

$$\begin{aligned} \sum_{i=1}^{k} f_i &= \sum_{i=1}^{k} e_i - \sum_{i=1}^{k} n_i + 2k \Rightarrow \\ \sum_{i=1}^{k} f_i &= e - n + 2k \end{aligned}$$

The external face is being counted k times in this sum, once for each component, so

$$\begin{aligned} f + (k-1) &= e - n + 2k \Leftrightarrow \\ f &= e - n + k + 1 \Leftrightarrow \\ f &\geq e - n + 2. \end{aligned}$$

Furthermore each edge contributes to only two faces, and each face is formed by at least three edges (assuming $e \geq 3$), so $e \geq \frac{3}{2} f$. Substituting, we get

$$\begin{aligned} f &\geq \frac{3}{2} f - n + 2 \Leftrightarrow \\ -\frac{1}{2} f &\geq -n + 2 \Leftrightarrow \\ f &\leq 2n - 4 \end{aligned}$$

(ii) Substituting $n = 6$ and $e = 12$ into $f \geq e - n + 2$, we find that $f \geq 8$. Suppose one of these faces has four or more edges. Then by drawing a new edge diagonally across it we can get another planar graph that has 13 edges and at least 9 faces. But this contradicts $e \geq \frac{3}{2} f$. ■

Exercise 581.

Suppose we want to construct a solid polyhedron using n pentagons (not necessarily regular), so that exactly three pentagons meet at a point. For what values of n is this feasible?

Solution.

This is an application of Euler's theorem on a closed surface ($f = e - v + 2$). We can think of such a polyhedron as a planar graph where lines between adjacent pentagons, and the faces are the pentagons themselves. Each vertex has exactly three edge endings, so $3v = 2e$. Each edge borders exactly two faces, so $2e = 5f$. Substituting into Euler's formula we get

$$f = \frac{5}{2}f - \frac{2}{3}\left(\frac{5}{2}f\right) + 2$$

Solving for f we obtain $f = 12$, so there must be exactly 12 polygons forming the polyhedron (as, for example, in a regular dodecahedron). ■

Exercise 582.

Choose n points on a circle so that no three chords meet at a point, and draw all possible chords. Use Euler's theorem to determine the number of regions into which the circle is divided by the chords.

Solution.

This is an application of Euler's theorem for planar graphs ($f = e - v + 2$). The number of regions of the circle is the number of interior faces of a graph, whose vertices are the points on the circle and the points of intersection inside the circle, and whose edges are the lines between vertices. For each set of four points outside the circle there is an additional point inside the circle where two chords cross. Thus there are n points on the circle and $n(n-1)(n-2)(n-3)/24$ points inside the circle; so $v = n + n(n-1)(n-2)(n-3)/24$. Each of the outside points has degree $n+1$ (for a chord going to each other point, and the sides of the circle), and each of the inside points has degree 4. Thus the total degree is $n(n+1) + n(n-1)(n-2)(n-3)/6$, and $e = n(n+1)/2 + n(n-1)(n-2)(n-3)/12$. Now,

$$\begin{aligned} f &= \frac{n(n+1)}{2} + \frac{n(n-1)(n-2)(n-3)}{12} - n - \frac{n(n-1)(n-2)(n-3)}{24} + 2 \\ &= \frac{n(n-1)}{2} + \frac{n(n-1)(n-2)(n-3)}{24} + 2 \end{aligned}$$

The answer is one less than this (since we do not include the outside face), which is

$$\frac{n(n-1)}{2} + \frac{n(n-1)(n-2)(n-3)}{24} + 1$$

■

Exercise 583.

The Doughnut Problem. Draw a K_6 graph on the surface of a doughnut (also known as *torus*).

Solution.
If you can do it (drawing it not solving it), please let me know! (My drawing skills are insufficient.) ■

Exercise 584.
Prove that in a connected planar graph where every vertex has degree at least three, there exists a region with fewer than six edges in its boundary.

Solution.
Assume this is not true (i.e. for some such graph every boundary has at least six edges). We use Euler's theorem ($f = e - v + 2$) to derive a contradiction. Since every vertexhas degree at least three, the total degree in the graph ($2e$) must be at least $3v$. Substituting $e \geq \frac{3}{2}v$ into Euler's theorem gives us $f \geq \frac{1}{2}v + 2$. Since every face has at least six edges, the total number of face borders ($2e$, since each edge borders two faces) must be at least $6f$. Substituting $e \geq 3f$ into Euler's theorem gives us $f \geq 3f - v + 2$, and hence $f \leq \frac{1}{2}v - 1$. Thus we have a contradiction. ■

8.7 Coloring on graphs

Exercise 585.
Given a graph $G = (V,E)$ a vertex coloring is an assignment of distinct colors to its vertices such that no pair of adjacent vertices have the same color.
Given a graph $G = (V,E)$ an edge coloring is an assignment of distinct colors to its edges such that no pair of edges incident on the same vertex have the same color.
The chromatic number $\gamma(G)$ (or $\chi(G)$) of G is the smallest number of colors needed in a vertex coloring.
The chromatic index $\chi(G)$ (or $\gamma(G)$) of G is the smallest number of colors needed in an edge coloring.
What are the γ, χ for K_4?

Solution.
For K_4, we have $\gamma(K_4) = 4$ and then $\chi(K_4) = 3$. ■

Exercise 586.
What are the γ, χ for K_{n_1,n_2}?

Solution.
Let $D = max_u d(u)$. For K_{n_1,n_2}, we have $\gamma(K_{n_1,n_2}) = 2$ and then $\chi(K_{n_1,n_2}) = D$.

■

Exercise 587.
Show that for generic graph G, we have $2 \leq \gamma(G) \leq D+1$ and then $D \leq \chi(G) = D+1$, where $D = max_u d(u)$.

Solution.
Omitted. ■

Exercise 588.

Kempe (1875). Every planar graph $G=(V,E)$ is 5-vertex colorable. Prove the result, if possible, by induction. We note that the following result is stronger.
Appel+Haken+Computer(1976). Every planar graph is 4-vertex colorable.

Solution.
The result will be proven by induction on the vertices of the planar graph $G=(V,E)$.
Before that we utilize Corollary 8.1.

> **Claim 8.1**
> In every planar graph there must be a vertex of degree 5 or less.

The proof is by way of Corollary 8.1. Otherwise every vertex has degree 6 or more. Given that the sum of the degrees is twice the number of the edges we can then show $2m=\sum_i d(i)\geq 6n$ which implies $m\geq 3n$ that would then contradict the $m\leq 3n-6$ of Corollary 8.1.
Base case ($n=1$). It is true for $n=1$ obviously.

Inductive step. Suppose it is true for all planara graphs with $k<n$ vertices. Consider a $k+1$ vertex planar graph G.
Case 1. Vertex v, $d(v)<5$. Pick a vertex v of degree $d(v)\leq 5$. If the degree of v is less than 5 i.e. $d(v)<5$ by induction, $G-v$ can be colored with 5 colors, and then v can be colored with a color not utilized by its four or fewer neighbors to provide a 5-coloring to G.
Case 2. Vertex v, $d(v)=5$. Suppose v has $d(v)=5$. Similarly as before, $G-v$ by induction can be colored with 5 colors. Let v_1,v_2,v_3,v_4,v_5 be the five vertices ordered according to the planar embedding of G, whose edges are incident on v. If Vertices $v_1,\ldots,v_5$ have fewer than 5 distinct colors altogether, G can become 5-colorable by the argument of the previous case.
In the remainder we discuss the case where vertices $v_1,\ldots,v_5$ have 5 distinct colors. Let without loss of generality the color of v_i be i. Let V_i be the set of vertices of $G-v$ with color i.

> **Claim 8.2**
> In $G_{13}=(V_1\cup V_3, E\cap V_1\times V_3)$, v_1 and v_3 are in the same connected component of G_{13}.

Solution. If they were not we could flip the color of one or the other in one connected component, and thus we would color v_1,v_3 with the same color, free one color, and use it to color v with the fred color. ■

For the same reason,

> **Claim 8.3**
> In $G_{14}=(V_1\cup V_4, E\cap V_1\times V_4)$, v_1 and v_4 are in the same connected component of G_{14}.

Likewise,

> **Claim 8.4**
> In $G_{25}=(V_2\cup V_5, E\cap V_2\times V_5)$, v_2 and v_5 are in the same connected component of G_{25}.

Because of the planarity of G, then the path $v_1\rightsquigarrow v_3$ would intersect or the path $v_1\rightsquigarrow v_4$ would intersect a path $v_2\rightsquigarrow v_5$ contradicting the planarity of G, unless one of v_2,v_5 are in a in a different component than the other and then we save up a color (2 or 5) to color v. ■

Exercise 589.
Show that for any regular degree three graph with a Hamiltonian cycle $\chi(G) = 3$. (Recall $\chi(G)$ and $\gamma(G)$ are the minimum numbers of colors used in edge and vertexcolorings resepectively of G).

Solution.
Since the graph is regular of degree 3, it has an even number of vertices, otherwise the sum of the degrees of the vertices would be odd. Therefore the Hamiltonian cycle is of even length and we can color its edges alternately using two colors, say 1 and 2. If we delete the edges of the Hamiltonian cycle from the original graph, we get a matching (we reduce the degree of each vertexby 2 by removing the cycle, and every vertexhad initially degree 3). We color the edges of the matching using a third color, say 3. ■

Exercise 590.
Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ (where V_1 and V_2 need not be disjoint). Define $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$. Show that

$$\gamma(G_1 \cup G_2) \leq \gamma(G_1)\gamma(G_2)$$

Solution.
Let α be a minimum coloring for G_1, i.e. a mapping that maps vertices of G_1 onto the $\gamma(G_1)$ colors we need to color G_1. Let β be the corresponding coloring for G_2. Then we color a vertex v of $G_1 \cup G_2$ using color $(\alpha(v), \beta(v))$ (the "colors" are ordered pairs here). If a vertexexists only in G_1 or in G_2 then we "fill" the other part of the pair with a color arbitrarily chosen from the set of colors used to color G_2 (or G_1 respectively). It is easy to see, that such a coloring is a valid one (since by its definition it doesn't color with the same color two vertices that are adjacent in $G_1 \cup G_2$ because that would imply that these two vertices which must be adjacent in G_1 or G_2 must also be colored the same color there, a contradiction to the validity of the initial coloring of G_1 and/or G_2) and the total number of colors used is equal to at most $\gamma(G_1) \cdot \gamma(G_2)$
You can show that equality holds in the case where G_1 is a circuit on 6 vertices and G_2 is $K_6 - G_1$. One can see that $\gamma(G_1) = 2$ and $\gamma(G_2) = 3$ while $\gamma(G_1 \cup G_2) = \gamma(K_6) = 6$. This also serves as a counterexample to an incorrect claim that $\gamma(G_1 \cup G_2) \leq \gamma(G_1) + \gamma(G_2)$. A counterexample to a claim $\gamma(G_1 \cup G_2) \leq \gamma(G_1) + \gamma(G_2) - 1$ is to have K_8 instead of K_6 above. ■

Exercise 591. Let K_n denote the n vertex complete graph. Show that

$$\chi(K_{2k}) = 2k-1$$

and

$$\chi(K_{2k+1}) = 2k+1$$

Solution.
(i) We first prove that $\chi(K_{2k}) = 2k-1$. We use color 1 to color the edges, where one of the $2k$ points is located in the center of a circle and the other $2k-1$ are on it. By rotating the edges one step clockwise around the center point, we get a new set of edges and we color them using color 2. We repeat the rotations until we color all the edges with colors from $1,\ldots,2k-1$. Each of these rotations uses a different set of edges and all the rotations exhaust all edges. From Vizing's theorem this is the best possible coloring we can find.
(ii) We start with the graph and rotate its edges around the center of the cycle (no point on the center now), to get all $2k+1$ possible colorings (note, that for each rotation, one vertex doesn't have an edge incident to it colored with the color used for the edges of this rotation). Now, we only have to prove that this number of colors is minimum, i.e. $\chi(K_{2k+1}) = 2k+1$. Using the same argument with that presented in class for K_5, if we try to color K_{2k+1} with only $2k$ colors, since the total number of edges is $k(2k+1)$, there must exist $k+1$ edges colored with the same color. This is a contradiction, since it implies a matching of size $k+1$ in a graph with $2k+1$ vertices. ■

Exercise 592.

If $G=(V,E)$ and $V_1 \subset V$, let $G(V_1)=(V_1, E\cap(V_1\times V_1))$.

(i) Show $\exists V_1, V_2 \subset V$ such that $V=V_1\cup V_2$ and $\gamma(G(V_1))+\gamma(G(V_2))=\gamma(G)$.
(ii) Show $\exists V_1, V_2 \subset V$ such that $V=V_1\cup V_2$, $V_1\cap V_2=\emptyset$, and $\gamma(G(V_1))+\gamma(G(V_2))>\gamma(G)$.
(Hint: Let $G(V_1)$ be a maximum complete subgraph of G).

Solution.
The first part is rather easy. Color G with $\gamma(G)$ colors. Let V_1 be the set of vertices of G colored with a particular color (among the $\gamma(G)$ colors), and V_2 be the rest of the vertices of G. Therefore, a coloring of G also colors $G(V_1), G(V_2)$, hence, we have that $\gamma(G(V_1))+\gamma(G(V_2))\leq\gamma(G)$. The $<$ cannot hold in the previous inequality because otherwise we could color G with less than $\gamma(G)$ colors (using the $\gamma(G(V_1))+\gamma(G(V_2))$ colors of the subgraphs, and coloring the vertices of G as they were colored in $G(V_1), G(V_2))$, a contradiction to the minimality of γ.
First note that, $\gamma(G(V_1))+\gamma(G(V_2))\geq\gamma(G)$ is always true since a valid coloring of $G(V_1)$ and $G(V_2)$ is also a valid one for G (assuming the colors used to color V_1 are different from those used for coloring the V_2 vertices). We would now prove that for $G(V_1)=K_k$, the maximum complete subgraph of G, equality cannot hold in the previous inequality. We prove it by contradiction. Let's assume we can color both $G(V_1)$, $G(V_2)$ with $\gamma(G)$ colors, i.e. $\gamma(G(V_1))=k$, $\gamma(G(V_2))=\gamma(G)-k$. Then, we would prove that we can eliminate one color from $G(V_1)$, and therefore color G with only $\gamma(G)-1$ colors, a contradiction to the minimality of $\gamma(G)$. Suppose, w.l.o.g., we color $G(V_1)$ with colors $1,\ldots,k$ and $G(V_2)$ with colors $k+1,\ldots,\gamma(G)$. Then, w.l.o.g., we would eliminate color 1 from $G(V_1)$ and color the vertex colored with color 1 (only one such vertex exists since $G(V_1)$ is a complete graph) with say (w.l.o.g.) color $\gamma(G)$.
Let v be the vertex colored with color 1 in $G(V_1)$. We then consider all the neighbors of v in G that belong to $G(V_2)$ and are colored with color $\gamma(G)$. Let these neighbors be $w_1,\ldots,w_l$. No two w vertices are connected with an edge since that would contradict to the identical coloring of these vertices. In graph theoretic terms, the set of vertices $w_1,\ldots,w_l$ is called *independent set*. Each of the w vertices is not connected to at least one of the vertices in $G(V_1)$ (otherwise, we could get a complete subgraph of size one more). Let w is not connected to u of $G(V_1)$ (in case we have more than one choices for coloring w we choose one color arbitrarily).
Then, we can color this w with the color of the vertex u in $G(V_1)$, and repeating this for all w vertices we can finally color all neighbors $w_1,\ldots,w_l$ of v (initially colored $\gamma(G)$) with colors taken from the set $\{1,\ldots,k\}$. Note, that some of the w might be colored with the same color, but this is admissible, since the w's form an independent set. Now we color v with $\gamma(G)$ since this color is available (we make it available for v by recoloring all its $\gamma(G)$-colored neighbors). Now, color 1, does not participate in the coloring of G (since only vertex v was colored 1 and now is colored $\gamma(G)$), and therefore we have a $\gamma(G)-1$ coloring of G, a contradiction. ■

Exercise 593.

An *independent set* is a set of vertices such that no two are connected by an edge. Let $i(G)$ be the maximum size of all independent sets in a graph G and $\gamma(G)$ be its chromatic number. Show that

$$i(G) \cdot \gamma(G) \geq n$$

where n is the number of vertices of G.

Solution.

In every coloring of G with $\gamma(G)$ colors, the set of vertices colored the same form an independent set. Let the colors used in a minimum coloring of G be $1, \ldots, \gamma(G)$. Let V_k be the set of vertices of G colored with color k. We have that $\sum_{k=1}^{\gamma(G)} \mid V_k \mid = n$ where n is the number of vertices of G and $\mid \cdot \mid$ represents the cardinality of a set. Since each V_k is an independent set, we get that $\mid V_k \mid \leq i(G)\ \forall k$. Therefore,

$$i(G) \cdot \gamma(G) \geq \sum_{k=1}^{\gamma(G)} \mid V_k \mid = n.$$

■

Exercise 594.
Show that there exist two graphs $G_1 = (V, E_1)$, $G_2 = (V, E_2)$ where $G_1 \cup G_2 = (V, E_1 \cup E_2) = K_{2k}$ (again, K_{2k} is the complete graph on $2k$ vertices), such that $\gamma(G_1) = 2$ and $\gamma(G_2) = k$.

Solution.
Take k vertices from K_{2k} and form a complete bipartite graph $K_{k,k}$ with the other k vertices (this complete bipartite graph has k^2 edges). We can color $K_{k,k}$ with two colors. The rest of K_{2k}, when we delete the edges of $K_{k,k}$, becomes disconnected and the two connected components of it are two complete graphs K_k. We can color them using k colors (use the same k colors to color both of them). We know that $\gamma(K_k) = k$, and therefore our two graph G_1, G_2 are: $G_1 = K_{k,k}$ and $G_2 = K_{2k} - K_{k,k}$. In another problem we presented an example where equality was obtained in $\gamma(G_1 \cup G_2) \leq \gamma(G_1) \cdot \gamma(G_2)$. ■

Exercise 595.

Show that the following inequalities hold for a graph G and its complement $\bar{G}$:

$$\gamma(G)+\gamma(\bar{G}) \leq n+1$$

$$n \leq \gamma(G)\gamma(\bar{G}) \ ,$$

where n is the number of vertices of G.

Solution.
We show that $\gamma(G)+\gamma(\bar{G}) \leq n+1$ by induction on n. A one vert exgraph and its complement both have chromatic number one, so base case is trivial. Now assume the relation holds for all graphs of size $n-1$ and consider an n vertex graph G. Let x be some arbitrary vertex of G, and let G' be the graph obtained by deleting x and its incident edges from G. By induction we know that $\gamma(G')+\gamma(\bar{G'}) \leq n$. Clearly $\gamma(G) \leq \gamma(G')+1$ and $\gamma(\bar{(}G)) \leq \gamma(\bar{G'})+1$ as we can always color x differently from all other vertices in G' or $\bar{G'}$ to obtain a legal coloring. If one of those inequalities is strict, then we have $\gamma(G)+\gamma(\bar{G}) \leq \gamma(G')+\gamma(\bar{G'})+1 \leq n+1$, and the relation holds. Thus the only remaining case is when both $\gamma(G)=\gamma(G')+1$ and $\gamma(\bar{G})=\gamma(\bar{G'})+1$. The former implies that the degree $d(x) \geq \gamma(G')$, since otherwise x cannot be adjacent to all colors in G', and may be colored with one already present. Similarly the latter implies $d(x) \leq (n-1)-\gamma(\bar{G'})$ as otherwise x would have fewer than $\gamma(\bar{G'})$ neighbors in $\bar{G'}$. These two inequalities imply

$$\gamma(G') \leq (n-1)-\gamma(\bar{G'})$$

Combined with the previous inequalities this yields

$$\gamma(G)+\gamma(\bar{G}) \leq (n-1)+2 = n+1$$

which proves the theorem.

To show $n \leq \gamma(G)\gamma(\bar{G})$, consider any coloring of G with $\gamma(G)$ colors. Clearly some color appears on at least $\frac{n}{\gamma(G)}$ vertices. None of these vertices are adjacent in G, so $\bar{G}$ contains the complete subgraph on these vertices. This implies

$$\gamma(\bar{G}) \geq \frac{n}{\gamma(G)}$$

from which the claim immediately follows. ■

Exercise 596.
Show that there exists a graph G which does not contain K_4 as a subgraph but requires at least 4 colors for a vertex coloring.

Solution.
An example is a wheel like graph. The outside cycle is of odd length and therefore requires three colors. A fourth is needed for the center point and its spikes. ■

Exercise 597.
Show whether the graph $G = (V, E)$ defined as follows is planar.

$$\begin{aligned} V &= \{x_0, \ldots, x_6, y_0, \ldots, y_6\} \\ E &= \cup_{0 \leq i \leq 6} \{(x_i, x_{i+1 \bmod 7}), (y_i, y_{i+1 \bmod 7}), (x_i, y_{7-2i \bmod 7})\} \end{aligned}$$

Solution.
This graph G is not planar, as demonstrated by an embedding of $K_{3,3}$ into G.

■

Exercise 598.
Show that if we color the points of the plane with 3 colors, there will be two points, 1 inch apart, with the same color.

Solution.
Consider four points as below that are the vertices of two equilateral triangles with one inch sides. If the claim is not true then point y must have the same color as point x (since points u and v must have the other two colors). Now consider rotating this figure around point x. All the points on the circle traced by the path of point y must be assigned the same color. But this circle as radius $\sqrt{3}$, and hence contains points that are an inch apart. ■

Exercise 599.
Show that there exists a graph that requires at least four colors for a vertex coloring but it does not contain K_4 as a subgraph.

Solution.
In the graph shown below (imagine that), the loop of length 5 requires 3 colors for any vertex coloring of it since it's of odd length. Therefore the central vertex requires a fourth color since it is connected to every vertex of the loop. By inspection the graph does not contain K_4 as a subgraph ■

Exercise 600.
Show that any n vertex graph G with more than $\frac{n^2}{4}$ edges has $\gamma(G) \geq 3$.

Solution.
Suppose that $\gamma(G) < 3$. Then, $\gamma(G)$ is either 1 (a trivial case that holds for trivial graphs with no edges, but we have at least one edge) or 2. For the latter case, $G = (V,E)$ is bipartite. Let $X \cup Y$ be a bipartition of the vertices in G. The maximum number of edges of G is then $|X||Y| = |X|(n-|X|)$. This quantity is maximized when $|X| = |Y| = n/2$, and therefore G can have at most $n^2/4$ edges, a contradiction (we have more than this number of edges). This proves that $\gamma(G) \geq 3$. ■

Exercise 601.
Show that for every graph G we have that $\gamma(G) \geq \frac{n^2}{n^2-2m}$ where, n, m are the number of vertices and edges of G. You may use, if you wish, Cauchy's inequality

$$(a_1^2 + \ldots + a_k^2)(b_1^2 + \ldots + b_k^2) \geq (a_1 b_1 + \ldots + a_k b_k)^2.$$

Solution.
Let $V_1, \ldots V_{\gamma(G)}$ be a partition of G's vertices according to some $\gamma(G)$ coloring of the graph. G can contain no edges from $V_i \times V_i$ for any of these subsets. Consider the $n \times n$ adjacency matrix for G. There are n^2 entries, all but $2m$ of which are 0 (consider the diagonal entries to be 0). These 0 entries include all those corresponding to edges between vertices from the same V_i. Thus,

$$|\, V_1 \,|^2 + \ldots + |\, V_{\gamma(G)} \,|^2 \leq n^2 - 2m$$

Cauchy's inequality tells us (chosing all the b_i's to be 1)

$$(|\, V_1 \,|^2 + \ldots + |\, V_{\gamma(G)} \,|^2)\gamma(G) \geq (|\, V_1 \,| + \ldots + |\, V_{\gamma(G)} \,|)^2 = n^2$$

Combining these two relations, we get

$$n^2 - 2m \geq \frac{n^2}{\gamma(G)}$$

from which the desired result follows. ■

8.8 Matchings

In the remainder of this section, all graphs are undirected (and simple).
The following definitions will be used.

Definition 8.33

For an undirected and simple graph (no self-loops and multiple edges) the following definitions are applicable.

- $\alpha(G)$ is the maximum independent set size of a graph G.
- $\beta(G)$ is the minimum vertex set cover size of a graph G.
- $\alpha'(G)$ is the maximum matching (edge) size of a graph G.
- $\beta'(G)$ is the minimum edge set cover size of a graph G.

Exercise 602.

Gallai ([17], 1959).

Let $G=(V,E)$ be an n vertex and m edge simple undirected graph with $|V|=n, |E|=m$. Let $\alpha(G)$ be the maximum independent set size of G. Let $\beta(G)$ be the minimum vertex set cover size of G. Then the following is true.

$$\alpha(G)+\beta(G)=n.$$

Solution.

We first show the following.

I is a maximum independent set if and only if $V-I$ is a minimum vertex set cover of G.

We first prove the forward direction.

$\Rightarrow$. I is a maximum independent set only if $V-I$ is a minimum vertex set cover of G.

Let I be a maximum independent set. Consider $V-I$. Pick an arbitrary edge $e=(a,b)\in E$ of G. It can't be that both $a,b\in I$ because then I could not be an inpependent set as it would then contain connected a and b and we know that $\forall x\in I\forall y\in I:(x,y)\notin E$. Therefore at least one of a,b belongs to $V-I$ and thus $V-I$ is a vertex cover.

We have shown that $V-I$ is a vertex cover. It remains to show that $V-I$ is a minimum vertex cover. Say, for the sake of contradiction, that $V-I$ is NOT a minimum vertex cover. Then $\exists u\in V-I$ such that $V-I-\{u\}$ to be indicated in the remainder as $V-I-u$ is also a vertex cover. Thus for every $e=(a,b)\in E$ at least one of a,b is in $V-I-u$. Consider the neighborhood of u and one of its edges, say edge $e_u=(u,c)\in E$. Because u is not in $V-I-u$ then c must be and thus c is also in $V-I$. This implies that all adjacent vertices to u are in $V-I$ and thus u is non-adjacent to vertices in I. Then $I+u$ is an independent set contradicting to the assumption that I is a maximum independent set. Thus $V-I$ is a minimum vertex cover.

$\Leftarrow$. I is a maximum independent set if $V-I$ is a minimum vertex set cover of G.

Let $V-I$ be a (mimimum) vertex cover for a given set I. Then for every $e=(a,b)\in E$ at least one of a,b must be in $V-I$. If both a,b were in I and not in $V-I$ then $V-I$ would not be a vertex cover.

Pick any two vertices of I for example $i_1\in I$ and $i_2\in I$. Then $e_{12}=(i_1,i_2)\notin E$, since otherwise $e_{12}\in E$ and by the prior statement for an arbitrary $e\in E$ we know that one of i_1,i_2 must be in $V-I$ but this would contradict the assumption that both are in I, not $V-I$. Thus $e_{12}=(i_1,i_2)\notin E$ for all $i_1\in I$ and for all $i_2\in I$ leads to the fact that I is an independent set.

We know show that I is a maximum independent set. Say I was a maximal but not maximum independent set. Then there exists an i such that $I+\{i\}$ is a larger independent set. If $I+i$ is an independent set then the following applies.

$$\forall i_j\in I: e_j=(i,i_j)\notin E. \tag{8.1}$$

Consider then an edge $e_i=(i,x)\in E$. From Eq. 8.1 it is clear that the following is true.

$$\forall x: e_i=(i,x)\in E\Rightarrow x\in V-I. \tag{8.2}$$

This is because $x\notin I$ and thus $x\in V-I$. If x were in I, it would contradict Eq. 8.1 (for $i_j=x$). Therefore $V-I-\{i\}$ is a cover of G because every edge without i as an end-point has an end point in $V-I$ (it is a vertex cover) and also in $V-I-\{i\}$, and every edge with end-point i has the other end-point in $V-I-\{i\}$. Therefore $V-I-\{i\}$ becomes a (smaller) vertex cover thus contradicting the minimality of $V-I$.

At the same time $|I|+|V-I|=n$. If I is a (maximum) independent set then $|I|=\alpha(G)$. Then $|I|+|V-I|=n$ leads to $|V-I|=n-\alpha(G)$. Furthermore $\beta(G)=\min_I|V-I|\leq|V-I|=n-\alpha(G)$, implies $\beta(G)\leq n-\alpha(G)$ and thus $\alpha(G)+\beta(G)\leq n$.

If $V-I$ is a (minimum) vertex cover then $|V-I|=\beta(G)$. $|I|+|V-I|=n$ leads to $n-|V-I|=n-\beta(G)=|I|\leq max_I|I|=\alpha(G)$, implying, $\alpha(G)+\beta(G)\geq n$. We conclude that $\alpha(G)+\beta(G)=n$, as needed. ■

Exercise 603.

Let M be a matching of a bipartite graph $G=(A\cup B,E)$. An alternating path of G relative to M is a set of edges alternating between edges of M end edges in $E-M$ (i.e. not in M).
Prove the following. Let M_1 and M_2 be two matchings of a bipartite graph $G=(A\cup B,E)$. Consider the subgraph of G defined as follows: $G=(A\cup B,M_1\oplus M_2)$ with edge-set $M_1\oplus M_2$, the symmetric difference of M_1 and M_2, that is, $M_1\oplus M_2=(M_1-M_2)\cup(M_2-M_1)$. Each connected component of this symmetric difference is one of the following types.

(1) An isolated vertex,

(2) A cycle of even length with edges alternating in M_1 and M_2,

(3) An alternating path with end-points distinct and both unsaturated in either one or the other matching.

Solution.
Let $a\in A$. We do a case analysis. Let $H=(A\cup B,M_1\cup M_2)$. In H a vertex a is of degree at most 2. A vertex in M_1 has degree one as it can be only in one edge of matching M_1; likewise for matching M_2. So in H, a can be of max degree 2.
A degree 0 vertex of $A\cup B$ is isolated (type 1). It is neither in M_1 nor M_2.
A degree 1 vertex belongs either to M_1 or M_2 since it can be in only one matched edge either of M_1 or M_2 in H. Moreover, it is the end-point of a path of M_1 and M_2 edges of H. (If it was an intermediate vertex it would be of degree 2.) Such a path has alternating edges from M_1 and M_2 (no vertex in a matching can be an end-point of two edges of it). Either the number of M_1 edges is equal to that of M_2 or one is one more than the other.
A degree 2 vertex can belong either to an alternating path of M_1 and M_2 edges discussed previously, or it belongs to a cycle of length even. (A cycle of H can not be of odd length; else two consecutive in the cycle edges of the same matching share a vertex which is impossible.)
We can do a case analysis as follows to describe the cases. $S(M_1)$ is the set of edges saturated by matching M_1. We can define similarly $S(M_2)$.
Case 1: $a\notin S(M_1-M_2)$ and $a\notin S(M_2-M_1)$. This means a is an isolated vertex.
Cases 2,3: $a\in S(M_1-M_2)$ and $a\notin S(M_2-M_1)$ and its symmetric case. This means a is in matching M_1 and moreover it is in M_1-M_2. There is no other edge in M_1-M_2 incident on a because a is in one matched edge. Moreover no edge of M_2-M_1 is incident on a either. Furthermore, $a\notin S(M_2)$ since otherwise a would be on an edge of M_2, but then it could not be $a\in S(M_1-M_2)$. Vertex a is of degree 1 then (type 3).
Case 4. $a\in S(M_1-M_2)$ and $a\in S(M_2-M_1)$ there exists an edge in M_1-M_2 incident on a and an edge in M_2-M_1 incident on a and no other edge is such. Vertex a is of degree 2 then and of type 2 (intermediate vertex of path) or of type 3 (a vertex of a cycle).
In case 1, a is of degree 0, in Cases 2 and 3 a is of degree 1 and in case 4, of degree 2. ■

Exercise 604.

Berge's theorem ([5], 1957). A matching M is maximum (i.e. of the longest size) if and only if there can't be an alternating path between two vertices that do not belong to $S(M)$. $S(M) = \{z | (z,b) \in M,$ **OR** $(a,z) \in M\}$.

Solution.

Only-If ($\Rightarrow$). Let M be a maximum matching and let us assume that an M-alternating path (i.e. an M-augmenting path) C exists with end points outside of the matching M. Then we can extend matching M using $M \oplus C$ to generate a matching of size one more than M contradicting the maximality of M. This is done by considering that the alternating path with end-points outside of M has k edges of M and $k+1$ edges outside of M i.e. in $C - M$. By unpicking the edges of M and picking the edges of $C - M$ (and also the edges of M not in the path) we generate a matching M' of one more edge than that of M, contradicting the maximality of M.

If ($\Leftarrow$). Let M be a matching. Let us assume that an alternating path DOES NOT exist between two vertices that do not belong to $S(M)$. Then we show that M is a maximum matching.

Say M (for the sake of contradiction) is not a maximum matching, and let M_1 be a maximum matching i.e. of size at least one more, ie. $|M_1| > |M|$. We know that for such a maximum matching M_1 the condition of the Only-if part is satisfied.

Consider $G_1 = (A \cup B, M \oplus M_1)$, where $M \oplus M_1$ is the symmetric difference of M and M_1 i.e. $M \oplus M_1 = (M - M_1) \cup (M_1 - M)$. In G_1 based on the previous theorem/problem, a vertex a is either isolated, or has degree 2 and is on a cycle of even length, or has degree 2 and is in a path as an intermediate vertex, or has degree 1 and is the end-point of a path. The first two cases are neutral with respect to the number of edges of M versus M_1. Only in the case of a path can we have either M or M_1 contributing one more edge if both end-points are from M or M_1 respectively. If an odd length path has both end-points in $M - M_1$ this is a violation of the Only-if part (matching M_1 can be extended by one). If an odd length path has both end-points in $M_1 - M$ this violates the assumption "an alternating path DOES NOT exist between two vertices that do not belong to $S(M)$". Thus M is a maximum matching. ■

Exercise 605.

Gallai ([17], 1959) and Norman and Rabin([31],1959), [8].

Let $G = (V,E)$ be an n vertex and m edge simple undirected graph with $|V| = n, |E| = m$. Let $\alpha'(G)$ be the maximum matching (edge set) size of G. Let $\beta'(G)$ be the minimum edge cover size of G. Then the following is true.

$$\alpha'(G) + \beta'(G) = n.$$

Solution.

Let M be a maximum matching of G. We then create C_1 as follows: we start from M and we add one edge $e = (x,y)$ per vertex y, where y is not in $S(M)$ and x (is expected to be in $S(M)$). Note $S(M)$ is the set of vertices saturated by M and thus $S(M) = 2\alpha'(G)$. Note also that it cannot be that both x,y are not in $S(M)$ as edge e could extend the maximum matching by one!

$$C_1 = M \cup \{e = (x,y) : x \in S(M), y \in V - S(M)\} \tag{8.3}$$

Note that C_1 by construction is an edge cover. Moreover we have the following.

$$|C_1| = |M| + (n - 2|M|) = n - |M|. \tag{8.4}$$

Let C be a minimum edge cover of G. We obtain a matching M_1 from C as follows. If a vertex u of C has degree two or more, we delete an incident edge say $e = (u,v)$. Moreover, v can have degree no more than one. This is so noting that edges of C cannot form a chain of length 3 such as a,b,c,d where $e_1 = (a,b)$ $e_2 = (b,c)$ and $e_3 = (c,d)$. If this was the case we could have removed e_2 from C and could have generated an edge cover C' smaller by one relative to the minimumc cover C, an impossibility (or contradiction to the minimality of C). Therefore the removal of an edge creates exactly one unsaturated vertex. We therefore conclude the following for the resulting matching M_1.

$$|C| - |M_1| = |V - S(M_1)| = n - 2|M_1| \tag{8.5}$$

Moreover,

$$|C| + |M_1| = n. \tag{8.6}$$

M_1 is a matching and M is a maximum matching. Therefore

$$|M_1| \leq |M|. \tag{8.7}$$

By way of Eq. 8.7 and Eq.8.6, Eq. 8.4 can be rewritten as follows,

$$|C_1| = n - |M| \leq n - |M_1| = |C| \tag{8.8}$$

But C is a minimal edge cover and therefore $|C| \leq |C_1|$. We then conclude that $|C| = |C_1|$. Then Eq. 8.4 with $|C| = |C_1|$ implies $|C| + |M| = n$ and using Eq. 8.6 $|C| + |M_1| = n$ we also conclude $|M| = |M_1|$. Either Eq. 8.6 or Eq. 8.4 yields the desired result:

$$\alpha'(G) + \beta'(G) = n.$$

■

Exercise 606.

P. Hall's theorem ([22], 1935).

Let $G = (A \cup B, E)$ be a bipartite graph, where $V = A \cup B$ and A and B define the bipartition of the vertices V of G. Then the following is true. Graph G has a (complete) matching M saturating A if and only if

$$\forall S \subseteq A : |N(S)| \geq |S|$$

where

$$N(S) = \{v \in B : \exists u \in S : (u, v) \in E\}.$$

If in addition $|A| = |B|$, the matching is a perfect matching.

Solution.

Only if. If G has a matching M, it must be $|N(S)| \geq |S|$ for all $S \subseteq A$. This is because if there is a $T \subseteq A$ such that $|N(T)| < |T|$, then there are not enough vertices of B in $N(T)$ to be matched by the A vertices of T.

If. Proof by induction. Assume that the result is true for all graphs with fewer than n vertices, i.e it is true for all $k < n$ vertices of A. We shall prove the following result.

$$\forall G, \quad \forall A, \forall k < n \quad |A| = k < n, \quad \textbf{if } \forall S \subseteq A, \quad |N(S)| \geq |S| \Rightarrow \textbf{ G has a (complete) matching}.$$

Base case is $n = 1$. True by inspection.

Inductive step.

Case 1 of inductive step : $\forall S \subseteq A, S \neq \emptyset, S \neq A : |N(S)| > |S|$**.**

Pick an edge $e = (a, b) \in E$ and remove edge e and also all other edges incident on a, b from graph G to get say graph $G' = G - a - b$. Graph G' has one fewer vertex in A after the removal of a (and also of b and of all those relevant edges). Furthermore, for all $S \subseteq A$, $N(S)$ has descreased possibly by one and thus $|N(S)| \geq |S|$. The remaining graph has one fewer vertex in A and one fewer in B and thus the induction hypothesis applies and a (complete) matching exists in G'. Such a matching saturates $A - \{a\}$. Adding e to this matching we get a matching of size one more that saturates all of A. Case 1 is settled.

Case 2 of inductive step : $\exists Z \subseteq A, Z \neq \emptyset, Z \neq A : |\mathbf{N}(\mathbf{Z})| = |\mathbf{Z}|$**.**

We split A into Z and $A - Z$. Z maps to $N(Z)$ and we have by assumption that $|N(Z)| = |Z|$. Moreover Z's and $A - Z$'s neighborhood is $N(Z)$ and $B - N(Z)$ respectively. Because $Z \neq A$, Z is at least one smaller than A, i.e. $|Z| < |A| = n$ or $|Z| \leq |A| - 1$.

Then the bipartite subgraph, $G_1 = (Z \cup N(Z), E \cap Z \times N(Z))$, with $|Z| < n$ satisfies the inductive hypothesis and a complete (and perfect) matching exists that saturates Z and its equal sized $N(Z)$.

The rest of the graph consists of $A - Z$ and $N(A - Z) \subseteq N(Z) \cup (B - N(Z))$. Since $Z \neq \emptyset$, $A - Z$ is such that $|A - Z| < |A|$ and thus the inductive hypothesis is potentially applicable. For this to be indeed the case, we need to dismiss the possibility that there exists an $A_2 \subseteq A - Z$ such that $|\mathbf{N}(\mathbf{A_2})| < |\mathbf{A_2}|$. Then the induction hypothesis would apply.

For the dismissal of that possibility we work as follows. We first observe that $Z \cup A_2$ has $|N(Z \cup A_2)| \leq |N(Z)| + |N(A_2)| < |Z| + |A_2|$. This contradicts the induction hypothesis $|N(S)| \geq |S|$ for all S including $S = Z \cup A_2$. Therefore an A_2 cannot exist with $|N(A_2)| < |A_2|$ i.e. the induction hypothesis applies. This generates a complete matching for $A - Z$ that if combined with the matching of Z leads to a complete matching for A. This completes the proof. ■

Exercise 607.

Kőnig ([24], 1931), Egerváry ([15],1931).

Kőnig proved the result for bipartite graphs; Egerváry proved the result for general graphs.
Let $G=(V,E)=(A\cup B,E)$ be a simple bipartite graph where $V=A\cup B$, $|V|=n$, $|E|=m$. Let $\beta(G)$ be the minimum vertex set cover size of a graph G. Let $\alpha'(G)$ be the maximum matching (edge) size of a graph G. Then the following is true.

$$\alpha'(G)=\beta(G).$$

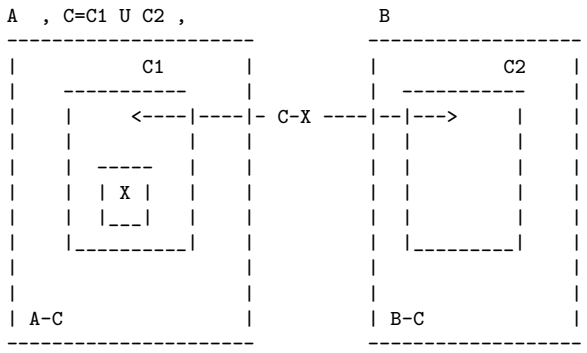
Solution.
We show first that $\beta(G)\geq\alpha'(G)$.
Consider a matching M. Every edge of it does not share a vertex with any other edge of it. Thus at least one vertex of a matched edge $e=(a,b)$ is part of any vertex cover C amd thus $|C|\geq|M|$ and then

$$\min|C|\geq\max|M|\Leftrightarrow\beta(G)\geq\alpha'(G).$$

We show next that $\beta(G)\leq\alpha'(G)$.
Consider a minimum vertex cover C of G of size $\beta(G)$. We then define two bipartite graphs as follows.

$$G_1=(A\cap C,B-C),\quad G_2=(A-C,B\cap C),\tag{8.9}$$

The figure below is given as a reference. $C_1=A\cap C$ and $C_2=B\cap C$. Therefore $C=C_1\cup C_2$ and $C_1\cap C_2=\emptyset$.



Graphs G_1 and G_2 cover between them all the edges and all the vertices of G. Moreover there is no edge from $A-C$ to $B-C$ since C is (minimum) vertex cover and thus any edge from $A-C$ or any edge from $B-C$ must be incident on C (either C_1 or C_2). We will use the existence or not of matchings in G_1 and G_2 respectively.

Let us assume that there is no matching in G_1 that covers $A\cap C$.
By Hall's theorem

$$\exists X,X\subseteq A\cap C:|N_{G_1}(X)|<|X|.$$

We claim and show the following.

Claim. Set $X'=(C-X)\cup N_{G_1}(X)$, where X as in Hall's theorem derived, is a vertex cover.
Consider $e=(u,v)\in E$. We distinguish two cases.

Case (i). Edge e has an end-point in $B\cap C$. Then edge e is covered indeed by X'.
Case (ii). Edge e has no end-point in $B\cap C$. Then one end-point is in $B-C$ and the other end-point must be in $A\cap C$ since C is a vertex cover. We focus our attention on $A\cap C$. There are two sub-cases in case (ii) then. First sub-case. If the end-point of e is not in X, where $X\subseteq A\cap C$, then it must in $C-X$, that is in $A\cap C-X$. This end-point is also covered by X'. Second sub-case. If the end-point of e is in X then the other end-point must be

in $N_{G_1}(X)$ which is part of X'. Thus e is also covered by X'.

Conclusion: Every edge e of E is covered by X' and thus the latter is a vertex cover of G. Moreover, the following applies, taking into consideration that for X, we have $|N_{G_1}(X)| < |X|$.

$$|X'| = |C| - |X| + |N_{G_1}(X)| < |C| - |X| + |X| = |C|$$

We thus found a cover X' smaller than the minimum vertex cover C, an impossibility (contradiction to the minimality of C).

Further conclusion: The assumption that there is no matching covering $A \cap C$ is false, i.e. there exists matching, call it M_1 that covers all of $A \cap C$.
With a similar and symmetric argument we work on G_2 to determine that there is a matching M_2 that covers $B \cap C$ respectively. Given that G_1 and G_2 are vertex disjoint, there is a matching $M = M_1 \cup M_2$ that has the following property.

$$|M| = |M_1| + |M_2| = |A \cap C| + |B \cap C| = |C|$$

Furthermore

$$\alpha'(G) = max|M| \geq |M| = |C| = \beta(G),$$

which yields $\alpha' \geq \beta(G)$. Combining this result and the first result shown we have the following.

$$\alpha'(G) = \beta(G).$$

■

Exercise 608.

(a) **Gallai ([17], 1959).** Let $G=(V,E)=(A\cup B,E)$ be an n vertex and m edge simple undirected bipartite graph with $|V|=n,|E|=m$, that is connected (and thus has no isolated vertices). Let $\alpha'(G)$ be the maximum matching (edge set) size of G. Let $\beta'(G)$ be the minimum edge cover size of G. Then the following is true.

$$\alpha'(G)+\beta'(G)=n.$$

(b) **Gallai ([17], 1959).** Moreover, in the absence of isolated vertices and assuming G has at least two vertices the following applies.

$$\alpha(G)=\beta'(G).$$

Solution.

Let M be a maximum matching. Therefore $\alpha'(G)=|M|$. From M we form C_1 as follows (see previous problem). Use vertex pool $P=V-S(M)$, i.e. the vertices not saturated by M. Pick a vertex u from P and add an edge incident on u to C_1. Set C_1 being formed and when this process gets completed is an edge cover. Moreover

$$\beta'(G)\leq|C_1|=|M|+(n-2|M|)=n-\alpha'(G).$$

Therefore

$$\alpha'(G)+\beta'(G)\leq n. \tag{8.10}$$

Let C be a minimum edge cover and $|C|=\beta'(G)$. Consider graph $G_1=(A\cup B,C)$. Since C is a mimimum edge cover

$$\forall e=(a,b)\in C\Rightarrow C-\{e\}\quad\text{is not an edge cover.}$$

Then edge e is such that either a or b (or both) are of degree one. (Note: they cannot be of degree zero because the graph is connected and has no isolated vertex, and if both were of degree two or more in C another edge would have covered both a,b and there would be no need for e in C contradicting the minimality of C.) Thus C has star like subgraphs. In order to form a matching M_1 pick an edge from each star. Then

$$\alpha'(G)\geq|M|=\text{NumConnComp}(G_1)=n-|C|=n-\beta'(G)$$

Therefore

$$\alpha'(G)+\beta'(G)\geq n. \tag{8.11}$$

From Eq. 8.10 and Eq. 8.11 we conclude the first result:

$$\alpha'(G)+\beta'(G)=n.$$

Furthermore, by way of

$$\alpha(G)+\beta(G)=n,$$

we conclude the following:

$$\alpha(G)+\beta(G)=\alpha'(G)+\beta'(G).$$

And from the Kőnig-Egerváry problem we have the following.

$$\alpha'(G)=\beta(G).$$

As a consequence, the second result follows:

$$\beta'(G)=\alpha(G).$$

■

Exercise 609.

The following problem is applicable to non-bipartite graphs as well.
An edge cover of a graph is a set of edges that touch every vertex. If M is a maximum matching, C is a minimum edge cover, and n is the number of vertices in the graph, show that $|M|+|C|=n$.

Solution.
Let $G=(V,E)$ a graph with $|V|=n$. Let M be a maximum matching. The vertices of G saturated by M are $S(M)$. We note that $|S(M)|=2|M|$. We get a cover C' by extending M. Each additional edge e added to C' saturates an unsaturated vertex of $V-S(M)$. Assuming a minimum cover exists, such edges e always exist; otherwise an edge (u,v) exists where $u\in V-S(M)$ and $v\in V-S(M)$ contradicting to the maximality of M (we could add then (u,v) to M extending its size by one).

$$C'=M\cup\{e=(a,b):e\in E,a\in S(M)\ ,b\notin S(M)\}$$

Each edge we add to the matching to extend a cover saturates one previously unsaturated vertex. Therefere the number of such edges we must add to M is equal to $n-2|M|$, The following is then derived.

$$|C'|=|M|+(n-2|M|)$$

and

$$|C'|+|M| \quad = \quad n \tag{8.12}$$

Let C be a minimum edge cover. We can derive a matching from C as follows (we give a 'high level' description). Remove edges from C which are incident to vertices of degree greater than 1, until each vertex would be of degree 1. In that way, we get a matching (each vertex is touched by at most 1 edge), and, let it be M'. Since paths of length three cannot exist in C (otherwise, if a path $v_1e_1v_2e_2v_3e_3v_4$ were part of C, then, we could get a smaller cover than C by dropping e_2 from C by using only edges $e_1,\ e_3$ which cover all 4 vertices v_1,v_2,v_3,v_4, a contradiction to the minimality of C. We then conclude that the removal of an edge from C creates exactly one unsaturated vertex. Therefore, we have:

$$|C|-|M'|=\ (\text{number of unsaturated vertices created})=n-2|M'|$$

which is equivalent to

$$|C|+|M'| \quad = \quad n \tag{8.13}$$

C is a minimum covering, therefore, $|C|\leq|C'|$ This implies by way of Eq. 8.12 and Eq. 8.13 that:

$$n-|C|\geq n-|C'|\Rightarrow|M'|\geq|M|$$

Since M is a maximum matching, we must have

$$|M|=|M'|$$

and again from Eq. 8.12 and Eq. 8.13 we get

$$|C|=|C'|$$

This means M,M' are both maximum matching and C,C' are minimum coverings of the same size. This implies that:

$$|M|+|C|=n$$

■

Exercise 610.

Frobenius ([20], 1917).

(a) Let $G = (A \cup B, E)$ be a bipartite graph. If $|A| = |B|$ and G is a k-regular graph, $k \geq 1$, then there exists a perfect matching involving all $a \in A$ and $b \in B$.

(b) Furthermore, show that the graph has k distinct perfect matchings.

Solution.

(a) If there is no perfect matchine, from Hall's theorem there exists an S such that $|N(S)| < |S|$. Each vertex of S has degree k. The set of edges of S that end (have the other end-point) in $N(S)$ is $|S|k$, and all such edges are distinct for a simple graph.

On the other side the $N(S)$ vertices of B are end-points of $|N(S)|k$ edges some of them originating from S and some of them possibly originating from outside S that is, $A - S$. The $|N(S)|k$ such edges must be at least equal to the $|S|k$ edges leading to $N(S)$. Therefore

$$|N(S)|k \geq |S|k$$

which leads to the contradiction $|N(S)| \geq |S|$ to $|N(S)| < |S|$. In other word the sum of degree argument establishes Hall's condition $|N(S)| \geq |S|$. By Hall's theorem the graph has a complete matching saturating A. Since $|A| = |B|$, this implies that the complete matching is indeed a perfect matching.

(b) For the second part, after a perfect matching is found in G, remove its edges from G. The resulting graph is a $(k-1)$-regular graph that also has a perfect matching. Continue likewise. ■

Exercise 611.

Frobenius ([20], 1917).
Let $G=(A\cup B,E)$ be a bipartite graph. Then G has a perfect matching if and only if $|A|=|B|$ and

$$\forall S\subseteq A: |N(S)|\geq |S|.$$

In the three problems to follow we prove the equivalence of the results that were given in the problem leading to this (or these) ones. This means the following equivalence will be established

$$P.Hall \Rightarrow \text{Kőnig} \Rightarrow P.Hall \Rightarrow Frobenius \Rightarrow \text{Kőnig}.$$

(a) Let $G=(A\cup B,E)$ be a bipartite graph, where $V=A\cup B$ and A and B define the bipartition of the vertices V of G. Show first the following

$$P.Hall \Rightarrow \text{Kőnig},$$

under the condition that the bipartite graph is connected and thus there are no isolated vertices.
(b) Use the same arguments slightly modified to show the following.

$$Frobenius \Rightarrow \text{Kőnig},$$

Solution.
(a) For bipartite graph G P. Hall's theorem takes the following form. Graph G has a (complete) matching M saturating A if and only if

$$\forall S\subseteq A: |N(S)|\geq |S|.$$

Case 1. If $|A|=|B|$ then the complete matching M, a derivative of Hall's theorem, is a perfect matching. In order to generate a minimum vertex cover of G we pick one vertex of each edge of the matching. Edges of the matching do not contain common vertices. The vertex cover C generated is a minimum cover. If none of the vertices of a matched edge e were picked and included into C, then edge e would not have been covered by C. Therefore $\alpha'(G)=\beta(G)$.
Case 2. If $|A|<|B|$ and let $|A|=p$ and $|B|=q$.
By Hall's theorem there is a complete matching M saturating A. This implies that A, the smaller in cardinality set between A and B is a cover of G, as G has no isolated vertices and thus $N(A)=B$. Thus a cover C of G is A i.e. $C=A$. This also implies $\alpha'(G)\leq\beta(G)$. Now let C be a minimum vertex cover of G and let $|C|=\beta(G)$. We are going to generate a matching M of the same size implying $\beta(G)\leq\alpha'(G)$ (that would then lead to $\alpha'(G)=\beta(G)$ by antisymmetry). The matching M to be generated is the union of two matchings M_1 and M_2. We show how to generate M_1. Form $G_1=(A\cap C,B-C)$. We first observe the following.

$$\forall S\subseteq A\cap C: |N(S)|\geq |S|.$$

Note first that $S\subseteq A\cap C$ implies $S\subseteq C$. If $|N(S)|\geq|S|$ were false we could replase C with $S\cup B\cap C$, and have created a new C_1 smaller in cardinality than C, a contradiction to the minimality of C. The Hall condition above for $S=A\cap C$ leads to $|A\cap C|\leq|B-C|$. By Hall's theorem we have a complete matching M_1 saturating $A\cap C$. The other M_2 would be generated by forming $G_2=(A-C,B\cap C)$, and lead to a complete matching saturating $B\cap C$. The union $M=M_1\cup M_2$ saturates C and is in size equal to the cardinality of C. Thus $\beta(G)\leq\alpha'(G)$, and consequently $\beta(G)=\alpha'(G)$.
Case 3. If $|A|>|B|$ swap the roles of A and B and it becomes Case 2.

(b) The second question is treated similarly. We note that Frobenius requires $|A|=|B|$. Therefore case 1 above still stands. For case 2 (and by symmetry for case 3), we need to modify our arguments. Form $G_1=(A\cap C,B-C)$. If $|A\cap C|<|B-C|$, we extend the cardinality of $A\cap C$ to match that of $B-C$ by adding dummy vertices and connecting them to all vertices of $B-C$. Then the Frobenius condition is applicable to generate M_1, now a perfect matching. Case 3 is handled similarly.

Alternative proof. showing directly that a matching saturating A is a minimum vertex cover. Note that if $|N(S)| < |S|$ no matching of cardinality $|A|$ exists. Therefore, $|N(S)| \geq |S|$ for all $S \subseteq A$. Then there exists a complete matching M that saturates A. Picking the vertices of A we generate a vertex cover of G as G is connected and has no isolated vertices. We show that A is a minimum vertex cover. Say A is not and a cover C exists $|C| < |A|$. Let $C_1 = A \cap C$, and $C_2 = B \cap C$, and thus $C = C_1 \cup C_2$. Obviously $N(A - C_1) \subseteq C_2$. This is because any edge $e = (a,b)$ of G that has an end-point a in $A - C_1$ it must have its other end-point b in C_2 else C would not cover edge e. Then, by $N(A - C_1) \subseteq C_2$ we have $|A - C_1| \leq |N(A - C_1)| \leq |C_2|$ The union of $A - C_1$ and C_1 is A. Therefore, $|A| = |C_1 \cup (A - C_1)| \leq |C_1| + |N(A - C_1)| \leq |C_1| + |C_2| = |C|$. Thus the cardinality of A is that of minimum vertex cover, i.e. A is minimum vertex cover. ■

Exercise 612.

Let $G=(A\cup B,E)$ be a bipartite graph, where $V=A\cup B$ and A and B define the bipartition of the vertices V of G. Show first the following

$$\text{Kőnig} \Rightarrow P.Hall$$

under the condition that the bipartite graph is connected and thus there are no isolated vertices.

Solution.
If Kőnig conditions apply and G has a complete matching it means A is minimum vertex cover. Then $\forall S\subseteq A$: $|N(S)|\geq |S|$ since otherwise we would not have a complete matching.
Assume now that Kőnig conditions apply the maximum matching is equal in size to the minimum vertex cover and in addition $\forall S\subseteq A : |N(S)|\geq |S|$, and let us try to show that there is a complete matching M saturating A.
Let C be a minimum vertex cover and thus $|C|=\beta(G)$. We otbain thus $|C|=\beta(G)\leq |A|$ since A is a vertex cover under the no isolated vertex assumption which is given. If $|C|=|A|$ since C is a ninimum vertex cover of G and by Kőnig there exists a maximum matching of the same size $|C|$ which is also $|A|$ it means that this matching would saturate A and thus be a complete matching. Case completed. Thus we examine the case $|C|<|A|$. Let (as before, in a previous problem) $C_1=A\cap C$, and $C_2=B\cap C$, and thus $C=C_1\cup C_2$. Obviously $N(A-C_1)\subseteq C_2$. This is because

$$|N(A-C_1)|\geq |A-C_1|=|A|-|A\cap C|=|A|-|C_1|>|C_2|.$$

The last inequality derives from the fact that $|C|=|C_1|+|C_2|$ and $|C|<|A|$. As previously noted an edge $e=(a,b)$ that has one end-point in $A-C_1$ must have the other end-point in C_2 else C does not cover e. But $|N(A-C_1)|>|C_2|$ allows for an edge $e_1=(c,d)$ where $e_1\in A-C_1$, and $d\notin C_2$ i.e. $d\in B-C_2$. Such an edge e_1 would not be covered by C. Thus $|C|<|A|$ is not possible, therefore $|C|\geq |A|$, and given that $|C|=\beta(G)\leq |A|$ we indeed conclude $|C|=|A|$. Thus the previous case completed is the only case. ■

Exercise 613.
Let $G=(A\cup B,E)$ be a bipartite graph, where $V=A\cup B$ and A and B define the bipartition of the vertices V of G. Show first the following

$$P.\ Hall \Rightarrow Frobenius$$

under the condition that the bipartite graph is connected and thus there are no isolated vertices.

Solution.
Frobenius is a special case of P. Hall. ■

Exercise 614.

A maximum matching of $G = (A \cup B, E)$ is a matching M of maximum cardinality. The deficiency $\delta(X)$ for a set $X \subseteq A$ is defined as

$$\delta(X) = |X| - |N(X)|.$$

The deficiency of graph G is defined as the maximum deficiency of a subset of A.

$$\delta(G) = \max_{X \subseteq A} \delta(X) = \max_{X \subseteq A} (|X| - |N(X)|) \geq \delta(X), \quad \forall X \subseteq A.$$

We have $\delta(G) \geq 0$ since $\delta(\emptyset) = 0$.

Show that in any bipartite graph $G = (A \cup B, E)$ the size of a maximum matching is $|A| - \delta(G)$.

Solution.
Let $X \subseteq A$ be the set determining the maximum deficiency i.e. $\delta(G) = \delta(X) = |X| - |N(X)|$. Thus $|N(X)| = |X| - \delta(X)$ and therefore at most $|X| - \delta(G)$ vertices in B forming $N(X)$ can be matched to a vertex in A that is X. Or equivalently $|X| - \delta(G)$ of the X vertices of X (a subset of A) are saturated by a matching. For a maximum matching M, where $\alpha' = |M|$, the following applies.

$$\alpha' \leq |A| - \delta(G).$$

We show that a matching M of size α' exists, where $\alpha' \geq |A| - \delta(G)$. Thefore $\alpha' = |A| - \delta(G)$. We add $\delta(G)$ new vertices in B and connect them to every vertex in A. We call the new augmented graph G_1. In the new graph for every $X \subseteq A$ we have (note that $N_G(X) = N(X)$ in the original graph G) the following.

$$|N_{G_1}(X)| = |N_G(X)| + \delta(G) = |X| - \delta(X) + \delta(G) \geq |X|.$$

since $\delta(G) \geq \delta(X)$. Thus by Hall's theorem G_1 has a complete (maximum) matching that saturates A. If we remove from that matching $\delta(G)$ edges/vertices mapped to the newly inserted vertices into B, we get a matching of size at least $|A| - \delta(G)$, i.e. $\alpha' \geq |A| - \delta(G)$. The two conditions $\alpha' \leq |A| - \delta(G)$ and $\alpha' \geq |A| - \delta(G)$ lead to $\alpha' = |A| - \delta(G)$. ■

Exercise 615.

We have a chessboard from which we have removed the top left square and the bottom right square. (A square is indicated by X.) We want to cover all the board (of 62 squares) with domino pieces. A domino piece is a rectangle that can cover horizontally or vertically two squares. Can we cover the incomplete square with 31 domino pieces?

```
  X X X X X X X
X X X X X X X X
X X X X X X X X
X X X X X X X X
X X X X X X X X
X X X X X X X X
X X X X X X X X
X X X X X X X
```

Solution.

```
  X X X X X X X       W B W B W B W
X X X X X X X X     W B W B W B W B
X X X X X X X X     B W B W B W B W
X X X X X X X X     W B W B W B W B
X X X X X X X X     B W B W B W B W
X X X X X X X X     W B W B W B W B
X X X X X X X X     B W B W B W B W
X X X X X X X       W B W B W B W
```

If we color the squares with color B and W for black and white that the chessboard is, the two missing squares are of the same type. A domino piece can only cover one B and one W. If there 30 of one type and 32 of another type they cannot be covered by 31 domino pieces.

For a graph-theoretic proof, build a bipartite graph with one set containing the B squares and the other the W squares. An edge determines whether two squares are next to each other in the North, Sourth, East, West direction. The problems becomes a matching problem. ■

Exercise 616.

Let G be an n vertex graph $G=(V,E)$ with no isolated vertices and with every vertex having degree at most d that is $d(u)\leq d, \forall u\in V$. If α' is the number of edges in a maximum matching for G, show that

$$\alpha' \geq \frac{n}{d+1}.$$

Solution.

Suppose M is a maximum matching for G of size α'. Let $X=S(M)$ be the vertices of G matched in M, and let $Y=V-S(M)$ be the remaining vertices of G unmatched by M. Moreover $|X|=2\alpha'$ and $|Y|=n-2\alpha'$. No two vertices in Y may have an edge between them, since otherwise adding that edge to M would extend M. If (u,v) is in M, and u has a neighbor $w\in Y$, then v cannot be adjacent to any vertex$z\neq w$ in Y. Otherwise there would be an M-augmenting path w,u,v,z which contradicts the maximality of M. Since there are no isolated vertices, each of the $n-2\alpha'$ vertices in Y is connected to some vertex in X. But each of the α' pairs of vertices in X can only have at most $d-1$ distinct neighbors in Y. Therefore we derive the following.

$$\begin{aligned} n-2\alpha' &\leq \alpha'(d-1) \Leftrightarrow \\ n &\leq \alpha'(d+1) \Leftrightarrow \\ \alpha' &\geq \frac{n}{d+1}. \end{aligned}$$

■

Exercise 617.

Show that any edge in a regular bipartite graph can be included in some perfect matching.

Solution.
Let $G = (A \cup B, E)$ where $E \subseteq A \times B$ be a regular degree k bipartite graph. By the Frobenius problem, every degree-k regular graph has a perfect matching. We repeat the Frobenius problem Corollary with respect to unveiling k perfect matchings. Find a perfect matching M for G and consider $G' = (A \cup B, E - M)$. This is also a regular bipartite graph, of degree $k-1$. So G' has a perfect matching which may be removed leaving a regular bipartite graph of degree $k-2$, and so on, until we are finally left with a graph that has no edges. Since every edge was removed as part of some perfect matching for G, this proves the claim. ■

Exercise 618.

Show that a regular degree k bipartite graph has at least $k!$ different perfect matchings.

Solution.
This can be shown by induction on $n = \mid A \mid$ in the statement: If $G = (A \cup B, E)$ is a bipartite graph such that $\forall v \in A, deg(v) \geq k$, and if G has a perfect matching, then G has at least $k!$ different perfect matchings. (We're actually claiming a slightly stronger result than required, since G need not be regular).
For $n = k$ (n's smallest possible value), we have a complete bipartite graph, and all of the conceivable $n! = k!$ matchings are legitimate. Now assume the statement is true for $n \leq n'$ and consider a case where $n = n' + 1$. Call a subset $X \subset A$ critical if $\mid N(X) \mid = \mid X \mid$. If A has no critical subsets, any edge may be put in the matching (k choices for $x_1 \in A$), and the graph remaining after deleting this edge, its end-points, and all of their edges will still satisfy the criterion of Hall's theorem and it will have a perfect matching. $\mid A - \{x_1\} \mid = n'$, and every vertex of this graph has degree at least $k - 1$, so by induction, we get the $k(k-1)! = k!$ bound. If A is critical, there must be a perfect matching consisting only of edges between X and $N(X)$ and between $A - X$ and $B - N(A)$. Any combination of two such matchings forms a perfect matching for G, and every perfect matching for G must be a combination of two such matchings. Matching X and $N(X)$ is a similar problem of smaller size, so by the inductive hypothesis there are at least $k!$ possible solutions. ■

Exercise 619.
Show that if the vertices of a bipartite graph $G = (A \cup B, E)$ have degree $\leq k$, then there exists a matching that saturates (touches) all vertices of degree k.

Solution.
Let $G = (A \cup B, E)$ be a bipartite graph, which has degree $\leq k$. We make G regular of degree k by adding extra edges and possibly extra vertices. Note, that by doing so, we add *no* edges incident to vertices that already have degree k. In that way, we get a k-regular graph. By the Corollary of Hall's theorem the new graph has a perfect matching. The edges of the perfect matching that are incident to vertices of degree k in G are edges in E. If we collect these edges we get a matching that touches all vertices of degree k in G. ■

Exercise 620.
Show that every bipartite graph $G = (A \cup B, E)$ has a matching that saturates all vertices of maximum degree (in A or B).

Solution.
(This is the previous problem rephrased.) Let $G = (A \cup B, E)$ be a bipartite graph, which has degree $\leq k$. We make G regular of degree k by adding extra edges and possibly extra vertices. Note, that by doing so, we add *no* edges incident to vertices that already have degree k. In that way, we get a k-regular graph. By the Corollary of Hall's theorem (multigraph case), the new graph has a perfect matching. The edges of the perfect matching that are incident to vertices of degree k in G are edges in E. If we collect these edges we get a matching that touches all vertices of degree k in G.

■

Exercise 621.

Show that if $G = (V, E)$ is a graph with $|V| = n$, and $n = 2r$ for some integer r, then if $d(x) > r$ for every vertex $x \in V$, then every edge in E is included in some perfect matching.

Solution.

Since the undirected graph has an even number of vertices a perfect matching is feasible. Suppose we want to show that a particular edge say (x, y) is in some particular perfect matching M. One edge of this matching will be edge (x, y). We pick the other edges as follows.

Each vertex of G originally had degree at least $r + 1$ where $n = 2r$. We remove from G the vertices x, y and all edges adjacent to x or y. The resulting graph G' has $n - 2 = 2r - 2$ vertices and each vertex in this graph has degree at least $r - 1$, since the degree of a vertex connected only to either x or y is decreased by 1 while the degree of vertices adjacent to both x and y is decreased by 2. For graph G' we have $2r - 2$ vertices and each one has degree at least $r - 1$, therefore, from Dirac's theorem G' has a Hamiltonian cycle. Since the cycle is of even length we get a matching on these $2r - 2$ vertices by picking every other edge of the Hamiltonian cycle. This gives a matching of size $r - 1$ which along with edge (x, y) yields a perfect matching M for G. Since edge (x, y) was chosen arbitrarily, the proof is complete. ■

Exercise 622.

Suppose $G=(V,E)$ is a graph with matchings M and M' where $|M'|\geq|M|+k$. In the graph $(V,(M-M')\cup(M'-M))$,
(i) What is a non-trivial lower bound on the number of alternating paths with both end points saturated by M'?

(ii) If M' is a maximum matching, what must be true of all the alternating paths ?

Solution.
Let symbol $\oplus$ denote the symmetric difference of two sets, i.e $M\oplus M'=(M-M')\cup(M'-M)$.
(i) Since M' has at least k matched edges more than M, and since isolated points, cycles and alternating paths of even length have the same number of matched edges of M and M' in $M\oplus M'$, then only alternating paths with both end points in either $S(M)$ or $S(M')$ may contribute one more edge in M and M' respectively. Let c and c' be the number of such paths respectively $(c,c'\geq 0)$. Since M' has at least k edges more than M we must have that $c'-c\geq k$ and therefore $c'\geq k$.
(ii) If M' is a maximum matching we can't get a matching of size at least one more than M' and so, among the alternating paths in $M\oplus M'$ we can't have alternating paths where both end-points are in $S(M)$, since otherwise by flipping the edges of that paths we can get a matching of size one more than M', a contradiction to the maximality of M'. ■

Exercise 623.

Suppose $G = (X \cup Y, E)$ is a bipartite graph where for all $x \in X$ $d(x) \geq k$, and for all $y \in Y$ $d(y) \leq 2k$. Show G has a matching of size at least $\frac{|X|}{2}$.

Solution.

Take a set A of X vertices. There are at least $|A|k$ edges leaving A and since every vertexof $R(A)$ must have degree at most $2k$, we must have that $|R(A)| \geq |A|/2$. We now take an extra copy of Y, let it be Y', and we add new edges between vertices in X and Y' so that the graph between X and Y' is an exact copy of G. In this way, for every set A of X we double the range of A in $Y \cup Y'$ and therefore $|R(A)| \geq 2\frac{|A|}{2} = |A|$. Therefore the resulting graph G' has a complete matching. Let y' be the image in Y' of vertex y of Y (i.e. we connect y' to X the same way we connected y to X in the original graph G). We get a complete matching of G', let it be M. If both y and y' are in $S(M)$ for some vertex y of Y, we delete from M the edge incident to y'. If y' is in $S(M)$ but not y, let the edge incident to y' be (x, y'). Then we delete this edge from M and we add to M edge (x, y). These actions are legal since if (x, y') is an edge in G' so is (x, y). We then repeat the above step for every vertexin $Y \cup Y'$ and the resulting matching M may be of size at least $|X|/2$ and this bound is achieved if the initial matching M contains as many pairs of image vertices y and y'as possible. ■

Exercise 624.

Two Latin squares are distinct if they differ at any position. Show that the number of distinct $n \times n$ Latin squares is at least $n!(n-1)!\ldots 2!$. (Hint: Prove that a regular degree k bipartite graph has at least $k!$ different perfect matchings).

Solution.

We first show how we can fill a Latin Square of order $n \times n$ by using a bipartite graph representation of the problem. We get a bipartite graph $G = (X \cup Y, E)$ where $|\,(|\,X) = |\,(|\,Y) = n$ and let each of the $x_i \in X$ corresponds to column i and y_j corresponds to number j and we would like to fill legally this Latin Sqaure with numbers 1 thru n. Initially we connect every vertexx_i of X to every vertexy_j of Y and get a complete bipartite graph, which is regular of degree n. We find a perfect matching in this graph (we know such a matching exists). Let it be M_1. If (x_i, y_j) is an edge in this matching, then we fill the $i-th$ column of the first row of the Latin Square with number j. We repeat this for every other matched edge and we therefore at the end of this step have filled the first row of the Latin Square. We delete the matching from the regular degree n graph and thus get a regular degree $n-1$ degree graph and repeat this procedure for the remaining rows.

If a regular k-degree graph has at least $k!$ perfect matchings (we prove that this holds below) then we can choose a matching and fill the first row of a Latin Square in at least $n!$ possible ways, the second row in $(n-1)!$ ways, and so on, so that we finally get the desired lower bound.

We prove now that a regular degree k simple bipartite graph $G(X \cup Y, E)$ has at least $k!$ perfect matchings. We will actually prove a weaker result. The proof resembles the proof of Hall's theorem. We prove the lower bound using simultaneous induction on k and $|\,X\,|$ on the statement 'if every vertex in X of G has degree at least k and G has a perfect matching, then G has at least $k!$ perfect matchings'. Suppose that for every set A such that $\emptyset \neq A \neq X$ we have $|\,R(A)\,| > |\,A\,|$. Pick any vertex in X, say without loss of generality vertex x. Since x has degree k, we have k choices among the edges incident to x, to add to a perfect matching. Let (x,y) be one of the edges . Then we delete from G the x,y vertices and all edges incident to them. The resulting graph is such that for every set B of $X - \{x\}$ we have $|\,R(B)\,| \geq |\,B\,|$ as in Hall's theorem, i.e. we get a complete matching for this graph and a perfect matching for the original graph which includes edge $\{x,y\}$. Every vertex in $X - \{x\}$ of the resulting graph has degree at least $k-1$ and from the inductive hypothesis has at least $(k-1)!$ complete matchings. Each of these matchings combined with one of the k choices of the edge $\{x,y\}$ give a complete matching for G i.e. we get that at least $k(k-1)! = k!$ perfect matchings exist.

Now if there exists a set A_0 such that $|\,A_0\,| = |\,R(A_0)\,|$, following the arguments in the proof of Hall's theorem we claim that there exists a complete matching between A_0 and $R(A_0)$ and between $X - A_0$ and $Y - R(A_0)$ and these two partial matching give a complete matching for G. From the inductive hypothesis, each vertex in A_0 has degree k and assuming the bipartite graph is simple (no multiple edges) we get at least $k!$ complete matchings between A_0 and $R(A_0)$ and therefore for G. ■

Exercise 625.
Suppose an $n \times n$ table has the first m rows, as well as the first entry in row $m+1$, filled with integers from 1 to n, where no number appears twice in any one row or column. Show that the table can be completed to form a Latin square.

Solution.
Consider just the first m rows of the table. As we saw in class this can be completed to form a Latin square. Such a Latin square must have some row from $m+1$ to n that has the same entry in column 1 as row $m+1$ of the table (since each number from 1 to n must appear once in column 1). If we switch that row with row $m+1$ in the Latin square, we get one that is a completion of the table. (Another way of solving this is observing that every edge of a regular bipartite graph is included in some perfect matching). ■

Exercise 626.

Suppose the entries in a $p \times q$ rectangle are filled in with integers from 1 to n ($p, q \leq n$) with no integer appearing more than once in any row or column. Show that this can be extended to an $n \times n$ Latin square if and only if each number appears in the $p \times q$ rectangle at least $p+q-n$ times.

Solution.

Wused the corollary to Hall's theorem for regular bipartite graphs to show that any legal setting for the first p rows can be extended to complete a Latin square. Thus we need to show that we can complete the remaining $n-q$ columns of the first p rows if and only if each number appears at least $p+q-n$ times in the first q columns of those rows. Suppose some number k appears fewer than $p+q-n$ times in the first q columns of the first p rows. In the remaining $n-q$ columns there may be at most $n-q$ more occurrences of k (otherwise two would be in the same column). Whatever we do, therefore, k must appear fewer than $(p+q-n)+(n-q) = p$ times in the first p rows, so the rectangle cannot be extended to a Latin square.

Otherwise, consider the bipartite graph

$$G = (X \cup Y, E), \quad X = \{1, \ldots, p\}, Y = \{1, \ldots, n\},$$

with

$$E = \{(i, j) : \text{the } i\text{'th row of therectangle does not contain the number } j\}.$$

Each vertexin X has exactly $n-q$ neighbors. Since every number appears in at least $p+q-n$ rows, every vertexin Y has at most $p-(p+q-n) = n-q$ neighbors. As shown in class we can edge color a bipartite graph of this degree with $n-q$ colors (call the colors $\{1, \ldots, n-q\}$). If edge (i, j) is colored k, we complete the first p rows by placing the number j in the entry at row i and column $q+k$. Note that each vertexin X must have exactly one incident edge from each color class, so this will indeed assign an entry to each remaining square in the first p rows. A number cannot be placed twice in the same column as this would imply it had two incident edges of the same color, nor can a number be placed twice in the same row with a valid coloring. ■

Exercise 627.
If $G = (X \cup Y, E)$ is a bipartite graph of degree Δ then for every $p \geq \Delta$ there exist p disjoint matchings $M_1, M_2, \ldots, M_p$ such that $E = M_1 \cup M_2 \cup \ldots \cup M_p$ and for $1 \leq i \leq p$ we have that:

$$\lfloor \frac{|E|}{p} \rfloor \leq |M_i| \leq \lceil \frac{|E|}{p} \rceil ,$$

where $\lfloor x \rfloor$ ($\lceil x \rceil$ respectively) is the largest (smallest) integer less (greater) than or equal to x.

Solution.
We showed that for a bipartite graph $\chi(G) \leq \Delta$. Thus we can color the edges of G with p colors, and each color class will form a matching (some possibly empty). Thus there do exist p disjoint matchings whose union is E. To show their sizes can be balanced, it suffices to prove the claim that if $M, N \subseteq E$ are two disjoint matchings with $|M| > |N|$ then there exist disjoint matchings M' and N' with $M' \cup N' = M \cup N$, but with $|M'| = |M| - 1$ and $|N'| = |N| + 1$. So if M has two or more edges than N, we can decrease the difference, and hence we can take any $M_1, \ldots, M_p$ and cause no pair differ by more than one.

Now let's prove the claim. Each component of $G' = (X \cup Y, M \cup N)$ is either an even cycle alternating in M and N, a path alternating in M and N, or an isolated vertex. If $|M| > |N|$, there must be an alternating path beginning and ending with edges from M. Swap all edges from M to N and vice versa along that path. ■

Exercise 628.

Give an example of a regular degree three graph that has no perfect matching.

Solution.
The graph shown below (imagine that) is a regular degree-3 graph on 16 vertices and does not have a perfect matching. The reason for the latter is quite simple. The graph consists of a central vertex and three symmetric components of 5 vertices each. For a perfect matching to exist, the fifth vertex of one of the 3 components must be matched to the central vertex (while the other four vertices can only be matched to vertices within the component). Then, the fifth vertices of the other two components can't be matched to each other or to any other vertex of the graph i.e no perfect matching exists in G. ■

Exercise 629.

Let $X = \{1, 2, \ldots, n\}$, and let X_r denote the set of all subsets of X of cardinality r. Use Hall's Theorem to prove that: (a) For $r < n/2$, there exists a 1-1 function $f_r : X_r \to X_{r+1}$, such that $A \subset f_r(A)$ for every $A \in X_r$.
(b) For $r > n/2$, there exists a 1-1 function $g_r : X_r \to X_{r-1}$, such that $g_r(A) \subset A$ for every $A \in X_r$.

Solution.
(a) We draw a bipartite graph $G = (\mathscr{X} \cup \mathscr{Y}, E)$ with $\mathscr{X} = X_r$ and $\mathscr{Y} = X_{r+1}$. For an $A \in \mathscr{X}$, and $B \in \mathscr{Y}$ we draw an edge from A to B iff $A \subseteq B$. The degree of every such A vertex is $n-r$ (we can increment A to a set of size $r+1$ by filling the extra position in $n-r$ ways) while the degree of every B vertex is $r+1$ (there are $r+1$ ways we can delete an element from a set of size $r+1$ to get a set of size r). We now get a set $S \subseteq \mathscr{X}$. Every element in S (S is a set of sets) is joined to $n-r$ elements in $R(S)$, and every element in $R(S)$ is joined to at most $r+1$ elements is S. Following the proof technique that proved that an r-regular bipartite graph has a perfect matching, we claim that the edges going out of S, which are $|S|(n-r)$ are among the ones going into $R(S)$ and the latter ones are at most $|R(S)|(r+1)$, and we thus get $|S|(n-r) \leq |R(S)|(r+1)$. We also have that $r < n/2$ (and therefore $(n-r) \leq (r+1)$, which finally shows that $|R(S)| \geq |S|$, i.e. Hall's theorem holds and G has a complete matching that saturates $\mathscr{X}$. If we map every element $A \in \mathscr{X}$ to its matched element $B \in \mathscr{Y}$, we get the desired 1-1 function f_r (every element in $\mathscr{X}$ is mapped to exactly one element in $\mathscr{Y}$ and no two elements in $\mathscr{X}$ are mapped to the same $\mathscr{Y}$ element).
(b) Part (b) is a direct consequence of part (a) when we replace each set by its complement with respect to X. ■

Exercise 630.

Show that in a bipartite graph $G = (X \cup Y, E)$ ($|\,Y\,| \geq |\,X\,|$), a necessary and sufficient condition for there to exist a matching that simultaneously saturates X and $B \subset Y$ is that:
(a) X can be matched into Y, and
(b) B can be matched into X.

Solution.
One direction is straightforward. If there exists a matching that simultaneously saturates X and B this matching matches X into Y and B is also matched into X. The other direction is interesting.
Let M_1 be a matching that saturate all the vertices in X, and M_2 a matching that saturates $B \subset Y$. If M_1 saturates all vertices in B, i.e. $B \subseteq S(M_1) \cap Y$, then we are done, M_1 is the desired matching. Suppose now that there exists a vertex $v \in B$ that is not saturated by M_1 (we will repeat the procedure to be described below, for all such vertices in B). Our initial candidate for a matching that saturates both X and B is M_1. We'll modify this matching to get a new one that saturates X, all previously saturated B vertices, plus v. We then take the **longest** alternating path that alternates between edges in $M_2 - M_1$ and edges in $M_1 - M_2$, starting from v with an edge from $M_2 - M_1$ (such an edge exists due to the existence of M_2). Let w be the other end-point of this path C. Vertex w can't be in X because otherwise we can extend C by picking the edge in M_1 that saturates w, a contradiction to the maximum length of C. It can't also be in B for a similar reason (we can extend it then thru an edge of M_2). The only case left is for w to be a point in $Y - B$ (and C is of even length then). In this case we flip the edges of the path C (so that M_2 edges become M_1 ones, and vice versa). In that way, we include in the new M_1 matching, vertex v (a previously unsaturated by M_1, B vertex) and drop w off the matching. All other vertices of C are still in $S(M_1)$ but thru different edges. We repeat this procedure for all other vertices in B still unsaturated by the new M_1 matching, and the M_1 we'll finally get will saturate both X and B. ■

Exercise 631.
Construct, for every $k \geq 1$, a graph with *a unique* perfect matching so that every vertex of this graph is of degree at least k.

Solution.

We use an inductive construction to build a graph G_k with a unique perfect matching, and where every vertexhas degree at least k. To do this we use two copies of G_{k-1} as subgraphs. (Note that for $k = 1$ we can form G_1 simply by joining two points with a single edge). For a larger k, build G_k as indicated in the drawing, where the new vertices x and y are connected to all the vertices in separate copies of G_{k-1}. Since G_{k-1} has a perfect matching it must have an even number of vertices. Thus the only way we can have a perfect matching in G_k is to match x with y. The existence and uniqueness of the matching follow from the uniqueness and existence of perfect matchings for G_{k-1}. ■

Exercise 632.

Consider the problem of covering an 8×8 chess board with dominos, where each domino occupies two adjacent squares along the same row or column, and no two dominos overlap. It is possible to completely cover the chess board with 32 dominos. If two diagonally opposite corner squares are removed, prove or disprove that the remaining 62 squares can be covered with 31 dominos.

Solution.
This problem is a variation of another similar problem. The two corners that are removed must be of the same color. So, if we remove them (and, say they are white squares) there are left 30 squares of one color (white) and 32 squares of the other color (black). Every domino piece covers, because of the way it is placed on the chessboard, one black and one white square. Since we have 32 black squares, there's no way we can cover them with 31 domino pieces. ■

Exercise 633.
Let $G=(V,E)$ be any bipartite graph and suppose $k \geq 1$. Show that G is the union of k edge disjoint spanning subgraphs $G_1,\ldots,G_k$ (i.e. $G_i=(V,E_i)$ with $E_i \subset E$, $E_i \cap E_j = \emptyset$ for $i \neq j$, and $\cup E_i = E$) such that

$$\left\lfloor \frac{d(x)}{k} \right\rfloor \quad \leq \quad d_{G_i}(x) \quad \leq \quad \left\lceil \frac{d(x)}{k} \right\rceil$$

(where $d(x)$ is the degree of vertex x in G, $d_{G_i}(x)$ is the degree of x in G_i, and $\lfloor\ \rfloor$ and $\lceil\ \rceil$ represent the floor and ceiling functions respectively).

Solution.
We draw a new graph G' from graph G. For every vertex x in the graph we split this vertex into $\lceil \frac{d(x)}{k} \rceil$ vertices, so that $\lfloor \frac{d(x)}{k} \rfloor$ of them have degree k and the last one has degree $d(x)-k\lfloor \frac{d(x)}{k} \rfloor$. Then every edge incident to x, touches now one of the new vertices. Every vertex of this new graph is of degree at most k, and from a very well known Corollary, the edges of this new graph is the union of k matchings (or $\chi(G')=k$). Now we identify the vertices of G' that correspond to the same vertex of G, and then, the k matchings, let them be $M_1,\ldots,M_k$, correspond to k subgraphs of G, call them $G_1,\ldots,G_k$. Since M_i is a matching, there is at most one edge of it incident to any of the $\lceil \frac{d(x)}{k} \rceil$ vertices of G' that correspond to x in G. Therefore $d_{G_i}(x) \leq \lceil \frac{d(x)}{k} \rceil$. On the other hand, there are $\lfloor \frac{d(x)}{k} \rfloor$ vertices of G' corresponding to x of G of degree k. There must be an M_i edge saturating each one of these $\lfloor \frac{d(x)}{k} \rfloor$ vertices, and hence, $d_{G_i}(x) \geq \lfloor \frac{d(x)}{k} \rfloor$. This proves the theorem, and G_i are the desired subgraphs. ■

Exercise 634.

Consider the problem of covering an 8×8 checker board with dominos, where each domino occupies two adjacent squares along the same row or column, and no two dominos overlap. It is possible to completely cover the checker board with 32 dominos. If two diagonally opposite corner squares are removed, prove or disprove that the remaining 62 squares can be covered with 31 dominos.

Solution.

This is not possible. Any placement of a domino will cover one black and one white square, but two opposite corner squares are the same color. Of the remaining 62 squares, 30 are one color and 32 are the other. No more than 30 dominos may be legally placed.

The graph representation of this problem is a bipartite graph where one side corresponds to the black squares, the other side corresponds to the white squares, and the edges correspond to adjacency between squares. The question then becomes whether this graph has a perfect matching. ■

Exercise 635.
There are rs couples at a dance. The men are divided into r groups, with s men in each group, according to their ages. The women are divided into r groups, with s women in each group, according to their heights. Show that r couples can be selected such that every age group and height group is represented.

Solution.
Consider the standard bipartite graph for the marriage problem where the men are on one side, the women on the other, and there are edges between the two members of a couple. Now condense the women into r super vertices, where each super vertexconsists of the s women of the same height. Do likewise for the men according to age. For each edge in the original graph there is an edge between the two super vertices containing the man and the woman of the original couple. The condensed graph is a regular degree s bipartite graph (as each super vertexhas s edges coming out of it). By the corollary to Hall's theorem it has a perfect matching (actually the condensed version may be a multigraph, but Hall's theorem still applies). The edges of such a perfect matching represent r couples with the desired property that each age and height is represented. ■

8.9 Adjacency matrix and list

Exercise 636.

Let $G=(V,E)$ has adjacency matrix A. Form the matrix products $A, A^2 = A*A, A^3 = A^2*A, \ldots, A^k$. Then $a_k(i,j)$ is the (i,j) entry of A^k. The following properties are shared by A^i, where $i=1,\ldots,k$. A i.e. $a(i,j)$ gives the number of paths of length 1 from a vertex i to a vertex j. A^2 i.e. $a_2(i,j)$ gives the number of paths of length 2 from a vertex i to a vertex j. A^k i.e. $a_k(i,j)$ gives the number of walks of length k from a vertex i to a vertex j. (In a walk, a vertex may get repeated.)

Solution.
Proof by induction. Base case is obvious. Let A^m shows the number of walks of length m from i to j.
We compute $A^{m+1} = A^m * A$. Note that

$$a_{m+1}(i,j) = \sum_{p=1}^{p=n} a_m(i,p) * a(p,j)$$

The first term of the sum $a_m(i,p)$ shows by induction the number of walks from i to a given vertex p of length k. Likewise $a(p,j)$ shows the number of walk from p to j of length 1. Thus $a_m(i,p)*a(p,j)$ gives the number of walk of length $m+1$ from i to j through penultimate vertex p. We repeat this for every candidate p and the total number of walks becomes $\sum_{p=1}^{p=n} a_m(i,p)*a(p,j)$. Result is shown. ■

Exercise 637.
Let $G=(V,E)$ has adjacency matrix A. For $B=A+A^2+A^3+\ldots+A^k$. Then $b(i,j)$ gives the number of walk of length at most k from i to j.
Let $G=(V,E)$ has adjacency matrix A. For $C=A+A^2+A^3+\ldots+A^{n-1}$. Then $c(i,j)$ gives the number of walk of length at most $n-1$ from i to j. Form T such that $T_{ij}=T(i,j)=1$ if $C(i,j)>0$ and $T_{ij}=0$ if $C(i,j)=0$. The matrix T is known as the transitive closure of graph $G=(V,E)$. The element $T_{ij}=T(i,j)$ is one if there is a walk from i to j, and zero if there is no walk from i to j.
Graph G is connected, or graph G is strongly connected if and only if T contains n^2 ones. (We assume $T_{ii}=1$ for all $1\leq i\leq n$ by default.)

Solution.
Omitted. ■

8.10 Ramsey numbers

Exercise 638.
Find the minimum number of vertices a graph must have to ensure there are at least two different 3-cliques or at least two different 3-independent sets, or a 3-clique and a 3-independent set. Two cliques (resp. independent sets) are different if they differ by at least one vertex.

Solution.
Six vertices are necessary and sufficient to guarantee a graph has two 3-independent sets and/or 3-cliques. It is easy to construct a five vertex graph without this property (e.g. consider a cycle). We now prove that six vertices are sufficient. Let $\{x_1,\ldots,x_6\}$ be those vertices. As shown in class there must be at least one 3-clique or 3-independent set. Without loss of generality assume that $\{x_1,x_2,x_3\}$ form a 3-clique. If there are edges between each pair in $\{x_4,x_5,x_6\}$ then those vertices form another 3-clique set, and we're done. Otherwise suppose (w.l.o.g.) there is no edge between x_4 and x_5. If any $y \in \{x_1,x_2,x_3\}$ is adjacent to neither x_4 nor x_5, then $\{y,x_4,x_5\}$ is a 3-independent set, and we are done. Otherwise assume each such y is adjacent to x_4 and/or x_5. Then either x_4 or x_5 (say the former w.l.o.g.) is adjacent to two vertices in $\{x_1,x_2,x_3\}$ (say x_1 and x_2). But then $\{x_1,x_2,x_4\}$ is a 3-clique. ■

Exercise 639.

A complete graph with n vertices denoted by K_n is an undirected graph where every vertex is connected to every other of the $n-1$ remaining vertices. The graph has $\binom{n}{2} = n(n-1)/2$ edges.
A k-klique is a graph that is isomorphic to K_k, the complete graph on k vertices. In other words, each vertex of the clique is connected to the remaining $k-1$ vertices of the clique.
A k-independent set is a graph whose complement is isomorphic to K_k. In other words, each vertex of the independent set is NOT connected to the remaining $k-1$ vertices of the clique. We shall call it k-is in short.
For any two positive integers p,q, there exists a minimum positive integer $R(p,q)$ that is known as the Ramsey number $R(p,q)$ such that for $r \geq R(p,q)$ if we arbitrarily color the edges of the complete graph with r vertices red or blue then we get

- either a red-complete graph with p vertices,
- or a blue-complete graph with q vertices.

Number r is noted as the $R(p,q)$ Ramsey number. [Another way to describe this problem is to say that that there is either a p-clique or a q-independent set.]
There is a party of two students. Either they know each other or they do not know each other. Show that $2 = R(2,2)$.
Consider $R(3,3)$. Show that $R(3,3) > 5$.

Solution.
Trivial and If you consider coloring the edges of K_5 you realize that $R(3,3) > 5$. ■

Exercise 640.

There is a party of 6 students. We claim that: either (i) some 3 people know each other (this is a 3-clique), or (ii) some 3 people do not know each other (this is a 3-independent-set).

Solution.
Let the 6 students be A,B,C,D,E,F.
Case 1. (A knows three or more). Let A know three others and let them be B,C,D. If two of the three friends of A know each other then A and those two know each other, (i) is true and we are done.
If none of them knows each other then B,C,D satisfy (ii) and we are done.
Case 2. (A knows two or less). Let A otherwise know two or less people and let those people be B,C. Then A does not know D,E,F. If D,E,F know each other we have found a triplet that satisfies (i) and we are done. If however there are two of D,E,F that do not know each other, and let those two be D,F, then consider the triplet A,D,F: none knows the other two and thus they satisfy (ii).
Consequently we proved $R(3,3) \leq 6$. Given (from the previous example) that $R(3,3) > 5$, we have $R(3,3) = 6$. ■

Exercise 641.

For $R(p,q)$ we expect $p>1$ and $q>1$, since a complete graph with one vertex has no edges. For all $p,q>0$ we have $R(p,q)=R(q,p)$. For all $k>0$ we have $R(1,k)=1$ and $R(2,k)=1$.
Show that $R(4,3)\leq 10$. That is in a 10-vertex graph either there is a 4-clique or a 3-independent set.

Solution.
Case 1: A knows at least 6 people. By the $R(3,3)=6$ result among the 6 there is either a 3-clique or a 3-is. In the first case the 3 plus A give a 4-clique. Done.
Case 2: A knows at most 5 people. That is A does not know 4 (total of $1+5+4=10$). Either all those 4 know each other and we found a 4-clique, or there are two who do not know each other. Those two plus A form a 3-is. Done. ■

Exercise 642.
Show that $R(4,4) = 18$.

Solution.
Omitted. ■

Exercise 643.

For $p \geq 2, q \geq 2$ we have

$$R(p,q) \leq R(p-1,q) + R(p,q-1)$$

Solution.

A. Use a people interpretation.

Consider a number of people. Among them pick 5. Call A the set of people who know 5. Call B the set of people who do not know 5.

We claim $R(p,q) \leq R(p-1,q) + R(p,q-1)$ and thus $c(A) + c(B) + 1 = R(p-1,q) + R(p,q-1)$.

- $c(A) \geq R(p-1,q)$**.** Set A of size at least $R(p-1,q)$ either contain $p-1$ people knowing each other or q people who do not know each other. The $p-1$ people who know each other also know 5 thus generating, including 5, a set of p people who know each other. Thus one way or the other for $A \cup 5$ we get p people knowing each other and for A we also have q people not knowing each other. Case completed.
- $c(B) \geq R(p,q-1)$ We then argue as follows for set B. Set B includes all people who do not know 5. B is of cardinality at least $R(p,q-1)$ and by induction either there exist p people in B who know each other or there are $q-1$ who do not know each other. In the latter case the addition of 5 generates q people not knowing each other. One way or the other the result has been proven.
- Otherwise $c(A) < R(p-1,q)$ and $c(B) < R(p,q-1)$ i.e. $c(A) \leq R(p-1,q) - 1$ and $c(B) \leq R(p,q-1) - 1$ thus giving $c(A) + c(B) + 1 \leq R(p-1,q) - 1 + R(p,q-1) - 1 + 1 \leq R(p-1,q) + R(p,q-1) - 1$. But this contradicts $c(A) + c(B) + 1 = R(p-1,q) + R(p,q-1)$ i.e. this cases IS NOT a case, it does not exist.

Thus $R(p,q) \leq R(p-1,q) + R(p,q-1)$.

B. We use a coloring of K_n interpretation.

Let $n = R(p-1,q) + R(q,p-1)$. We claim that $R(p,q) \leq R(p-1,q) + R(q,p-1)$.

Consider random coloring the edges of K_n with two colors, red and blue. Pick an arbitrary vertex of K_n and call it 5. Among the $n-1$ edges incident on 5 P are red and Q are blue where $P + Q = n-1$ and thus $P + Q + 1 = R(p-1,q) + R(q,p-1)$.

We do a case analysis

Case 1: $P \geq R(p-1,q)$**.** Then the red edges's end-points either contain a red K_{p-1} or a blue K_q. We are done in the latter case; in the former case adding 5 we turn the red K_{p-1} into a red K_p. Thus one way or the other a red K_p or a blue K_q is generated.

Case 2: $Q \geq R(p,q-1)$**.** The the blue edges' end-points either contain a red K_p or a blue K_{q-1}. In the former case we are done; in the latter case adding 5 to the $q-1$ blue edge end-points, we generate a blue K_q. Thus one way or the other a red K_p or a blue K_q is generated. ■

Exercise 644.

Show that $R(p,q)$ is finite for all $p > 1$ and $q > 1$.

Solution.
Proof is by induction of $p+q$.
Base case. The base case is for $p = q = 1$. Either a red or a blue 1-subgraph results and the statement is true by inspection trivially.
Inductive step. Now suppose that $R(p-1,q)$ and $R(p,q-1$ are finite. We proved before that $R(p,q) \leq R(p-1,q)+R(p,q-1)$. Thus $R(p,q)$ is finite as it is bounded by the sum of two finite numbers. ■

Exercise 645.
Show that $R(p,q) \leq \binom{p+q-2}{p-1}$.

Solution.
It follows by induction from $R(p,q) \leq R(p-1,q)+R(p,q-1)$ and simple algebraic manipulations. ■

Exercise 646.
Show that $R(k,k) \leq \binom{2k-2}{k-1} \leq 2^{2k}$.

Solution.
Omitted. ■

Exercise 647.
Show that for any $k > 4$, we have $R(k,k) \geq 2^{k/2}$.

Solution.
The number of n vertex graphs is $2^{\binom{n}{2}}$.
The number of n vertex graphs containing at least one k-clique is at most $\binom{n}{k} \cdot 2^{\binom{n}{2}-\binom{k}{2}}$.
The number of n vertex graphs containing either at least one k-clique or at least one k-is is at most $2 \cdot \binom{n}{k} \cdot 2^{\binom{n}{2}-\binom{k}{2}}$.
Thus if $2 \cdot \binom{n}{k} \cdot 2^{\binom{n}{2}-\binom{k}{2}} < 2^{\binom{n}{2}}$, then there is some n vertex graph that has neither a k-clique nor k-inpdendent set.
For this we need to show

$$2 \cdot \binom{n}{k} < 2^{\binom{k}{2}}$$

or equivalently

$$2 \cdot n^k / k! < 2^{k(k-1)/2}$$

and for $n = 2^{k/2}$ it suffices to show that

$$k!/(2 \cdot 2^{k/2}) > 1$$

which is true for $k \geq 4$. ■

Exercise 648.

Schur's Theorem. For any finite coloring of the positive integers there exist x,y,z such that $x+y=z$ and all three are of the same color.

Solution.
Color the numbers $1,...,S(r)$ with different colors. $S(r)=R(3,3,\ldots,3)$ i.e. the Ramsey number of order 3 and multiplicity r. Consider $K_{S(r)}$. Color edge (a,b) with $|b-a|$'s color. By definition of Ramsey numbers There exists a triangle $x>y>z$ whose edges are of the same color. Then $A=x-y,B=y-zC=x-z$ have the same color. Note that $A+B=x-y+y-z=x-z=C$. ■

Exercise 649.
Show that

$$r(k,k) \geq \frac{k}{e\sqrt{2}} 2^{\frac{k}{2}},$$

where $e = 2.71\ldots$.

Solution.
The number of n vertex labeled graphs with a k-clique (which by symmetry is the same as the number of n vertex graphs with a k-independent set) is at most

$$\binom{n}{k} 2^{\binom{n}{2}-\binom{k}{2}}.$$

This comes from the observation that for each possible k-clique, there are $\binom{n}{2} - \binom{k}{2}$ ways to complete the graph. To prove a lower bound for $r(k,k)$ it suffices to show that for smaller values of n,

$$\binom{n}{k} 2^{\binom{n}{2}-\binom{k}{2}} < \frac{1}{2} 2^{\binom{n}{2}}.$$

If fewer than half the graphs have a k-clique, and fewer than half have a k-independent set, then some must have neither.

That relation is equivalent to

$$\binom{n}{k} 2^{-\binom{k}{2}} < \frac{1}{2},$$

which is certainly true if

$$\frac{n^k}{k!} 2^{-\binom{k}{2}} < \frac{1}{2}.$$

Substituting for n and expanding k choose 2, the left hand side becomes

$$\frac{k^k 2^{\frac{k^2}{2}}}{e^k 2^{\frac{k}{2}} k!} \cdot \frac{2^{\frac{k}{2}}}{2^{\frac{k^2}{2}}} = \frac{k^k}{e^k k!}$$

By Stirling's formula we know

$$k! \geq (\frac{k}{e})^k \sqrt{2\pi k} e^{\frac{1}{12k+1}} \geq (\frac{k}{e})^k \sqrt{2\pi k}$$

And so the left hand side is at most $1/\sqrt{2\pi k}$, which is indeed less than $\frac{1}{2}$ for $k \geq 1$.

■

Exercise 650.

We give here the definition of the generalized Ramsey number $r_l(k_1,\ldots,k_l)$ as the smallest value of n so that if we arbitrarily color the edges of the complete graph K_n with l colors (say $\{1,\ldots,l\}$) then for some index i there exists a complete subgraph of size k_i whose edges are all colored with color i.
Show that for $r_l(3) = r_l(3,\ldots,3)$ (i.e. $k_i = 3\ \forall i \in \{1,\ldots,l\}$) we have that:

$$r_l(3) \leq e \cdot l! + 1\ ,\quad e = 2.7182\ldots$$

Solution.
We use induction on l. The result trivially holds for $l = 1$ and for $l = 2$ we use the result given in class for $r(k,k)$ in the base case $k = 3$. Now suppose that it holds for all values of l up to $l-1$. We are then going to prove that $r_l(3) \leq el! + 1$. Take the complete graph on $n+1 = \lfloor el! \rfloor + 1$ vertices. Pick a point x. This points has n edges incident to it. Since

$$\lfloor el! \rfloor = \lfloor \sum_{i=0}^{\infty} \frac{l!}{i!} \rfloor = \sum_{i=0}^{l} \frac{l!}{i!} = 1 + l \sum_{i=0}^{l-1} \frac{(l-1)!}{i!} = 1 + l \lfloor e(l-1)! \rfloor,$$

We got the third term from the second one because the $l+1$ first terms of the sum are integers and the missing terms, if one uses the remainder R_l, (which gives an upper bound on the sum of all terms from the $i = l+1$-st one to infinity) of the Taylor expansion of $el!$ sum up to at most $\frac{e^{\theta}}{l+1} < 1\ ,\ 0 \leq \theta \leq 1$. We thus get that among the n edges incident to x there must exist a set of $\lfloor e(l-1)! \rfloor + 1$ edges colored the same color, say, without loss of generality color l. Let the other end points of all these edges incident to x form set X. If there exists an edge, among the edges between the points in X, colored l, a triangle colored l is found. Otherwise, all the edges in the subgraph on the points of X are colored with $1,\ldots,l-1$ colors and the size of X is $\lfloor e(l-1)! \rfloor + 1$ and therefore from our inductive hypothesis we are done. This gives that indeed $r_l(3) \leq \lfloor el! \rfloor + 1$. ■

8.11 Random Graphs

Exercise 651.

Show that there exists a tournament on n vertices that has no transitive subtournament of size $\lceil 2\log n+1\rceil$.

Solution.
Let T be the class of all tournaments on n vertices, i.e. $\mid T \mid = 2^{\binom{n}{2}}$ and let T_s be the class of subsets of T which contain a transitive subtournament of size s. Then

$$T_s = \cup_{A\subseteq\{1,\ldots,n\}|A|=s} \cup_{\sigma \text{ an ordering over } A} \tau_{A,\sigma}$$

It is clear that $\mid \tau_{A,\sigma} \mid = 2^{\binom{n}{2}-\binom{s}{2}}$. Then, $\mid T_s \mid \leq \binom{n}{s} s! 2^{\binom{n}{2}-\binom{s}{2}}$ where the first term gives the number of ways of choosing s out of n points, the second the number of orderings (permutations) on s points (and each of them gives a transitive tournament on s vertices) and the last term is just $\tau_{A,\sigma}$ for fixed A, σ. Then,

$$\frac{\mid T_s \mid}{\mid T \mid} \leq \binom{n}{s} s! 2^{-\binom{s}{2}} < \frac{n^s}{s!} s! 2^{-\binom{s}{2}} = n^s 2^{-s(s-1)/2}$$

The last term is smaller than 1 if

$$n^s 2^{-s(s-1)/2} \leq 1 \;\Rightarrow\; s \geq 2\log n + 1$$

or equivalently $s \geq \lceil 2\log n+1\rceil$. Therefore, even for the smallest possible value of s we can find a tournament with no transitive subtournament of that size. ■

8.12 Spanning trees

Exercise 652.

Given a procedure to find a minimum cost spanning tree of a graph, show how one can instead find a spanning tree that has minimum weight among those spanning trees containing a fixed given forest as a subgraph.

Solution.
Let $G = (V,E)$, and let $V_1,\ldots,V_f$ be the connected components of the forest $F \subseteq E$ (possibly including isolated vertices). Contract G to form a new graph G' as follows.

$$G' = (V',E'), \quad V' = \{V_1,\ldots V_f\}, \quad E' = \{(V_i,V_j)\,|\,V_i \times V_j \cap E \neq \emptyset\}$$

Thus the vertices of V' are the connected components of F. An edge (V_i,V_j) implies an edge from a $u \in V_i$ to a $v \in V_j$, and $c'(u,v) = \min_{u,v} c(u,v)$, where $c(u,v)$ is the costs of edge (u,v) in G. Now just run the minimum cost spanning tree algorithm on G' and combine its ouput with the forest to form a spanning tree for G in the obvious manner. That is if the MST includes edge (V_i,V_j), the spanning tree for G will include the cheapest edge between a vertex in V_i and a vertex in V_j.
Clearly the output contins F as a subgraph. And it will be connected (as each component of F is conected by edges from G') and have no cycles (as there can only be one path between components of F), and hence it will be a spanning tree. Furthermore any spanning tree for G containing F can be constructed in this manner from a spanning tree for G', and the cost is just a constant amount more (the weight of F). Thus this spanning tree is minimum. ■

Part V

Probability

Chapter 9

Probability and random variables

9.1 Short review on probability theory

9.1.1 Experiments and trials

An experiment can be considered a sequence of individual trials. An experiment or a trial have outcomes. Outcomes form a sample space.

A trial might consist of rolling a die (plural: dice). There are six possible outcomes forming the sample space S of this trial (or elementary experiment). (The terms roll, throw, or toss might be used interchangeably.)

An experiment can consist of repeating for example the trial above of rolling a die three times. Or we can view it as a single trial involving the rolling of a die thrice. There are $6 \times 6 \times 6$ possible outcomes that form the sample space S of this experiment.

If we roll two dice at the same time the sample space has 6×6 possible outcomes.

Definition 9.1 **Experiment or trial**

An experiment (or trial) is any procedure that can be repeated and generates a well-defined set of outcomes.

If the experiment or trial is random (or stochastic), then we can call it a random experiment or random trial or just a random process. The terms experiment, or trial or random process are to be used interchangeably.

Definition 9.2 **Sample Space**

The set of all possible outcomes of an experiment is known as sample space. We denote a sample space with the symbol S.

Thus for the die experiment, $S = \{(a,b) : 1 \leq a,b \leq 6\}$. Note that in bibliography Ω is used to indicate a sample space.

Definition 9.3 **Sample point, outcome**

An element of S, an outcome of an experiment, is also known as a sample point.

That is, an outcome of an experiment is an element of the sample space S, a sample point. An event T is a subset of the sample space, and thus a set of outcomes. An event outcome is a success if it belongs to T; otherwise it is a failure. We may also call an event outcome favorable or unfavorable. If $T = \emptyset$ we call this an impossible event. If $|T| = 1$ it is a simple event. The complement of T is T' or T^C.

9.2 Events

Definition 9.4 **Event**

An event is a set of outcomes and is a subset of S.

Sometimes we refer to the outcomes (sample points) of S as elementary events.

Let us flip a coin three times. The sample space now is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\},$$

where H indicates a Heads outcome, and T a tails outcome. The sample space is of cardinality 2^n, where n is the number of repetitions of the coin experiment. An event $E_1 = \{HHT, HTH, THH\}$ indicates an event in which only one Tails was encountered in the experiment. (One can also describe this event as the event where the number of Heads is an even positive integer!) There are three possible outcomes of S associated with this event. Another event $E_2 = \{HHH\}$ is a no Tails event.

Definition 9.5 **Mutually exclusive events**

Two events A, B are mutually exclusive if and only if $A \cap B = \emptyset$. Three events or more are mutually exclusive if every two of them are mutually exclusive.

Continuing our most recent example, obviously E_1, E_2 are mutually exclusive events since $E_1 \cap E_2 = \emptyset$. We can now provide some formal definitions, including that of a probability space.

9.2.1 Event space and measurable spaces

Definition 9.6 **Event space**

An event space T is a set of events that is, a collection of subsets of a sample space S.

An event space T is usually a σ-algebra, also known as a σ-field.

9.2.2 σ-algebras

Definition 9.7 **σ-algebra or σ-field**

T is a σ-algebra or σ-field if it contains a collection of subsets of a sample space S that satisfy the following.

1. $S \in T$,
2. $A \in T$ implies that $A' = S - A \in T$ (closedness under complement), and
3. for a countable sequence $A_1, A_2, \ldots, A_i$, where $A_i \in T$ implies $\cup_i A_i \in T$ (closedness under union).

Using DeMorgan's Law $((A \cap B)' = A' \cup B', or (A \cup B)' = A' \cap B')$, one can show also that for a countable sequence A_i, $A_i \in T$ implies $\cap_i A_i \in T$ (closedness under intersection).

9.2.3 Measurable spaces

Definition 9.8 **(S,T): measurable space**

A pair (S, T) is known as a measurable space: it consists of a sample space S of outcomes, and a set of events T that is a σ-algebra.

The pair (S, T) is called measurable space because it is possible to put a measure on it. A σ-algebra looks similar to a topology. We provide a definition of a topology for informational purposes.

9.2.4 Topologies

Definition 9.9 **Topology**

Y is a topology (or topological space), if it contains a collection of subsets of a space S that satisfy the following.

1. $S \in Y$, $\emptyset \in Y$,
2. for an arbitrary sequence A_i, $A_i \in Y$ implies $\cup_i A_i \in Y$ (closedness under union),
3. for a finite sequence A_i, $A_i \in Y$ implies $\cap_i A_i \in Y$ (closedness under intersection).

Thus in a σ-algebra we have a countable sequence of unions (and by implication intersections) plus complementation, whereas in a topology we have an arbitrary sequence of unions and a finite sequence of intersections. The magic words are countable and arbitrary and finite.

A σ-algebra requires the complement of a set to be in T; a topology does not require it, though this is the case for the two sets S and $\emptyset$. A topology is associated with closeness (neighborhood) whereas a σ-algebra with length (measure or weight). With a topology we intend to determine whether a function is continuous or not. In probability spaces (σ-algebra) we want to assign values to sets (event probabilities). For every topology one can associate it with a σ-algebra to it through the notion of a Borel-space or algebra.

Example 9.2.1

A (S,T) is a topological space where $S = \{a,b,c\}$ and $T = \{\emptyset, S, \{a,b\}, \{b,c\}, \{b\}\}$. But it is not a σ-algebra since the complement of $\{a,b\}$ is not in T for example.

Example 9.2.2

A finite intersection of open intervals $\cap_{m\in\{1,3,5\}}(a-\frac{1}{m}, b+\frac{1}{m})$ preserves the open interval notion. On the other hand, a countable intersection of intervals $\cap_{m\in N}(a-\frac{1}{m}, b+\frac{1}{m})$ is a closed interval $[a,b]$. Consider also the case $a=b=0$ or $a=0, b=1$.

A σ-algebra is ready to be measured. This is why (S,T) is also known as a measurable space. Every element A of T can be assigned a number, a probability.

For the discrete case A has a number of elementary events (outcomes) of S and thus $P(A)$ or $P(\{A\}) = \sum_{s\in A} p(s)$, where $p(s)$ is the probability of an elementary event. For the continuous case A is an open or closed set and then the Lebesque measure of A becomes relevant. The Lebesque measure of the closed set (interval) $[a,b]$ is $b-a$ but so are of the open sets $[a,b)$ or $(a,b]$ or (a,b). An open set is a set that is not closed.

For a Borel σ-algebra we start with a topological space and then we include the minimal number of open intervals that can describe the Borel σ-algebra's T under countable unions, intersections and complements. Then an element of T for the corresponding Borel σ-algebra (or Borel algebra) is also known as a Borel set. For a sample space S of a measurable space (S,T), and $s \in S$, we call $\{s\}$ an elementary event, as already noted. Moreover $\emptyset$ and S are also events of T. The impossible event or null event is the $\emptyset$. If A, B are events so are $A \cap B$ and $A \cup B$ or A' for example.

A measure m is a non-negative function on $\mathbb{R}$ that can be applied to elements of the event space T of (S,T).

9.2.5 Measures

Definition 9.10 **Measure m**

A measure m is a function on T defined as $m : T \mapsto \mathbb{R}$ that satisfies the following.

1. $m(A) \geq m(\emptyset) = 0$ for $\forall A \in T$,
2. if $A_i \in T$ is a countable sequence of pairwise disjoint sets, $m(\cup_i A_i) = \sum_i m(A_i)$.

9.3 Probability space

A probability measure denoted by P is a measure were $\mathbb{R}$ is replaced by the closed interval $[0,1]$ and has one additional property: $P(S)=1$. Several times T is not well-established or defined. Then $T=\mathbb{P}(S)$, is the powerset of S.

9.3.1 Probability measure

Definition 9.11 **Probability measure P**

A probability measure P is a function on T defined as $P:T\mapsto[0,1]$ that satisfies the following.

1. $P(A)\geq P(\emptyset)=0$ for $\forall A\in T$, where $P(A)=\sum_{s\in A}P(s)$,
2. if $A_i\in T$ is a countable sequence of pairwise disjoint sets, $P(\cup_i A_i)=\sum_i P(A_i)$, and
3. $P(S)=1$. (Moreover, $P(A)\leq 1\ \forall A\in T$.)

The three properties (conditions) above are also known as Kolmogorov's Axioms. They are sometimes referred to as non-negativity, (countable) additivity and normalization properties.
Since $\emptyset$ and S are pairwise disjoint sets, by property (2) we have $P(S)+P(\emptyset)=P(S)$ and thus $P(\emptyset)=0$ cited in (1) can be obtained. For an event A, its complement A' is also an event because a $\sigma-algebra$ such as (S,T) is closed under complementation. For any event A and its complement event A', we have that A and A' are pairwise disjoint and $A\cup A'=S$. By property (3) we have $P(S)=1$ and by property (2) $1=P(S)=P(A\cup A')=P(A)+P(A')$. Utilizing property (1) we can infer $P(A)\leq 1$.
Several times when we use the term probability we mean probability space and we thus imply the existence of the triplet (S,T,P). So several times the existence of (S,T,P) is implicily provided.
The probability of an event is the fraction of the number of times/ways the event can occut over the number of possible outcomes. If we roll a dice, and define an event to be the roll of the dice turns an even number, the probability of this event is $3/6$ since three of the six outcomes map to odd numbers $\{1,3,5\}$ and the number of possible outcomes is six $\{1,2,3,4,5,6\}$.

9.3.2 Probability space

Definition 9.12 **Probability space**

A probability space is a triplet (S,T,P), where

- S is a sample space that is, a set of outcomes,
- $T\subseteq 2^S$ is a σ-algebra on S, that is, a collection of subsets containing S and closed under complement, closed under union (of a countable number of sets), and by DeMorgan implication closed under intersection (of a countable number of sets), and
- P is a probability measure on T.

A probability space is the triplet (S,T,P). Absence of any reference to T, we use $T=\mathbb{P}(S)$, or sometimes T can be omitted.

The set S is known as the sample space and the elements of S are known as outcomes or elemetary events. For an event $A\in 2^S$, we define $P(A)$ the probability of event A. The probability of an event A is the sum of the probabilities of the elementary events of A. If S is finite, (S,T,P) is a finite probability space. If S is finite and $T=2^S$ the probability measure is determined by its values on elementary events. Thus the probability space assigns through $p:S\to[0,1]$ a probability $p(s)$ to every element of s of S such that $p(s)\geq 0$, with $\sum_{s\in S}p(s)=1$.

Then for an event $A \in 2^S$, the probability of the event A is the sum of the probabilities of the elements of S in A i.e. the sum of the probabilities of the elementary events, or in other words $P(A) = \sum_{s \in A} p(s)$.

Theorem 9.1

Let (S,T,P) be a probability space. The following apply to events, $A,B,\ldots$ of T.

1. *Monotonicity: If $A \subseteq B$, then $P(A) \leq P(B)$.*
2. *Subadditivity: If $A \subseteq \cup_i A_i$, then $P(A) \leq \sum_i P(A_i)$.*
3. *Continuity below: If $A_1 \subset A_2 \subset \ldots$ and $\cup_i A_i = A$ then $\lim_{i\to\infty} P(A_i) = P(A)$.*
4. *Continuity above: If $A_1 \supset A_2 \supset \ldots$ and $\cap_i A_i = A$ with $P(A_i) < \infty$, then $\lim_{i\to\infty} P(A_i) = P(A)$.*

Definition 9.13 **Discrete probability space**

A discrete probability space has a sample space S that is a countable set. Then

$$P(A) = \sum_{s \in A} p(s),$$

where $P(s) \geq 0$ and $\sum_{s\in S} P(s) = 1$.

Sometimes we refer to it as finite probability space.

Proposition 9.1

If an event A contains a finite number of outcomes, say $A = \{a_1,\ldots,a_k\}$ then

$$P(A) = \sum_{a_i \in A} P(a_i) = \sum_i P(a_i) = P(a_1) + P(a_2) + \ldots + P(a_n).$$

One can prove this proposition by defining disjoint events $A_i = \{a_i\}$ and then use the second Kolmogorov axiom. Moreover if the sample space S is finite and then we denote the sample points s_i we have $\sum_i P(\{s_i\}) = P(\cup\{s_i\}) = P(S)$ and by the third Kolmogorov axiom $P(S) = 1$ concludes that $\sum_i P(\{s_i\}) = 1$.

Definition 9.14 **Equiprobable space**

For a finite probability space S of n sample points, if each sample point has the same probability as any other one, the sample space is called equiprobable space.

We provide the following definition.

Definition 9.15 **High probability even**

We say an event E dependent on n occurs with high probability, if $P(E) \geq 1 - 1/n^c$ for some constant $c > 1$.

Instead of writing "with high probability" we shall write "w.h.p." or "whp" instead.

9.4 Probability of events

Lemma 9.1

For any collection of events $A_1, \ldots, A_n$,

$$P(A_1 \cup \ldots \cup A_n) \leq \sum_i P(A_i).$$

Proof. Consider, events B_i such that

$$B_i = A_i - (A_1 \cup \ldots \cup A_{i-1})$$

Then $\cup_i B_i = \cup A_i$ and $P(B_i) \leq P(A_i)$ and the events B_i are disjoint. By additivity of the probability measure we have

$$P(A_1 \cup \ldots \cup A_n) = P(B_1 \cup \ldots \cup B_n) = \sum_i P(B_i) \leq \sum_i P(A_i)$$

■

Lemma 9.2 **Independent Events**

Two events A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

Theorem 9.2 **Properties of events**

Consider two events A, B of probability space (S, T, P).

$$P(\emptyset) = 0,$$

$$P(A - B) = P(A) - P(A \cap B),$$

and for $A \subseteq B$ we have

$$P(A) \leq P(B).$$

Moreover,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Example 9.4.1

A coin is tossed twice. The number of heads is recorded. The sample space is $S = \{0, 1, 2\}$. The following probability model is assigned $p(0) = 1/4, p(1) = 1/2, p(2) = 1/4$. Event $A = \{1, 2\}$ with $p(A) = 3/4$, and event $B = \{2\}$ with $p(B) = 1/4$,

Example 9.4.2 **Uniform distribution**

$P(A) = c(A)/c(S)$ for all $A \subseteq S$. Note $c(A)$ is the cardinality of set A also denoted $|A|$.

9.5 Conditional Probabilities

Theorem 9.3 **Conditional Probability**

Let A,B be two events of a finite probability space (S,T,P) and $P(B)>0$. The probability that an event A occurs conditional on event B had already occurred is known as the conditional probability of A given B and denoted $P(A|B)$. It is given by

$$P(A|B)=P(A\cap B)/P(B).$$

Moreover

$$P(A\cap B)=P(A|B)P(B).$$

Theorem 9.4 **Generalization of conditional probability**

$$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cap A_2)\cdot P(A_4|A_1\cap A_2\cap A_3)\cdot\ldots\cdot P(A_n|A_1\cap A_2\cap A_3\cap\ldots\cap A_{n-1})$$

Definition 9.16 **Properties of independent events**

Let A,B be two events of finite probability space (S,T,P). A is independent of B if

$$\textbf{A is independent of B}\ \Leftrightarrow P(A)=P(A|B).$$

Corollary 9.1

From $P(A|B)=P(A\cap B)/P(B)$ since $P(A|B)=P(A)$ we have $P(A)=P(A\cap B)/P(B)$ which implies that $P(A\cap B)=P(A)P(B)$. Moreover $P(B|A)=P(B)$.

Example 9.5.1

A (pair of) dice is tossed. The probability of any elementary event (outcome) of S is $1/36$. Let B be the event the sum of the dice is 6. What is $P(B)$? Let A be the event a dice is 5. What is $P(A|B)$? This reads probability the event A occurs given that B has already occurred. This is because $P(A\cap B)=2/36$. This is because $P(B)=5/36$. And $P(A|B)=P(A\cap B)/P(B)=(2/36)/(5/36)=2/5$.

Proof. We show B below.

$$B=\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

The probability that a dice is 5 if the sum is 6 is $2/5$. Then $A|B$

$$A|B=\{(1,5),(5,1)\}$$

Obviously $P(A|B)=2/5$. This is because $P(A\cap B)=2/36$, and $P(B)=5/36$. And $P(A|B)=P(A\cap B)/P(B)=(2/36)/(5/36)=2/5$. ■

Definition 9.17 **Independent Repeated Experiments**

Let (S,T,P) be a finite probability space. A space of n independent trials is space S_n consisting of ordered n-tuples of elements of S with the probability of an n-tuple defined to be the product of the probabilities of its components.

$$P((s_1,\ldots,s_n))=P(s_1)\ldots P(s_n).$$

9.6 Random variables

The sample space S of an experiment or trial or process has outcomes that might be number or might not be numbers. Think about the experiment of tossing a coin. Sometimes instead of using symbolic values or names to an outcome such as H or T we prefer to use numeric values such as 1 and 0 respectively. This mapping is known as random variable.

As we mentioned earlier, a probability space is a triplet (S,T,P). In the simplest of the cases, S is a finite set, and $T = 2^S = P(S)$, that is T is the powerset of S, the set of all posible subsets of S and thus T has cardinality $|T| = 2^{|S|}$. The probability measure P, $P : T \mapsto [0,1]$, essentially becomes, $P : S \mapsto [0,1]$, i.e. it is defined on the elementary events, and by extension to any event subset of T. Then

$$P(A) = \sum_{w \in A} P(\{w\}),$$

and

$$\sum_{w \in S} P(\{w\}) = 1.$$

Definition 9.18 **Random variable: simple definition**

A random variable (later, r.v.) X defined on a probability space (S,T,P) is a real-valued measurable function on S. Random variable X is then a mapping

$$X : S \mapsto \mathbb{R}.$$

Definition 9.19 **Random variable: more detailed definition**

A random variable (later, r.v.) X defined on a probability space (S,T,P) is a real-valued measurable function that assigns a real value $X(s)$ to an elementary event s of S so that for every x, where $-\infty < x < \infty$, set $\{s : X(s) \leq x\}$ is an event contained in T.

For a random variable we can informally write $P(X = 5)$ as a simplication of $P(\{X = 5\})$ to be defined formally as $P(\{s \in S : X(s) = 5\}$. Moreover, when we informally write $P(5 \leq X \leq 11)$ we actually mean $P(\{s \in S : 5 \leq X(s) \leq 11\})$.

Definition 9.20 **Convention**

For a random variable X, when we write $P(X = x)$ we mean $P(\{X = x\})$. Furthermore, we also mean

$$P(\{X = x\}) = P(X = x) = P(\{s \in S : X(s) = x\} = P(X^{-1}(\{x\})).$$

For a random variable X, when we write $X \in A$ for $A \subseteq \mathbb{R}$, we mean $\{s \in S : X(s) \in A\}$. Then,

$$P(X \in A) = P(\{s \in S : X(s) \in A\}).$$

9.7 Discrete random variables

Definition 9.21 **Discrete r.v.**

A random variable is discrete if its domain consists of a finite or a countably infinite set of outcomes S (sample space).

Definition 9.22 **Range**

The range $R(S,X)$ of random variable X is the set of numeric values assigned to outcomes of a sample space S by random variable X, i.e.

$$R(S,X) = \{X(s) : \forall s \in S\}$$

Random variable X is a discrete random variable if its range $Z = R(X,S)$ is countable.

Definition 9.23 **Discrete r.v.:formal definition**

Let (S,T,P) be a probability space and let Z be a set. A Z-valued discrete random variable on (S,T,P) is a function

$$X : S \mapsto Z,$$

where $X(S) = \{X(s) : \forall s \in S\} = Z$ is countable.

Example 9.7.1

In an experiment a six-faceted die is rolled. Let X detone the outcome of this expriment. The sample space $S = \{1,2,3,4,5,6\}$ and $X : S \mapsto \mathbb{Z}$, where $X(1) = 1$. An event is a subset of S; for example event E is the event that the die is rolled event, i.e $E = \{2,4,6\}$.

We usually denote an event involving random variable X by writing $X = x$. This indicates event $E = \{s \in S : X(s) = x\}$. Moreover, it should be the case that $E \subseteq T$.

Example 9.7.2

Flip a coin twice. The probability space defined is (S,T,P). $S = \{H,T\}^2 = \{H,T\} \times \{H,T\}$, $|S| = 4$ and for every elementary event (outcome) of S, say $s = (a,b) \in S$ we have $P(s) = 1/4$. Let X be a r.v. representing the number of head outcomes obtained.

$$X : S \mapsto \{0,1,2\},$$

where $X((a,b))$ is the number of heads in the two flips a,b, where a records the outcome of the first flip, and b of the second. If we had denoted an H with 1 and a T with a 0, then $X((a,b))$ would have been more easily defined as $X((a,b)) = a+b$.
(a) In order to calculate the probability of having exactly two heads we have the following.

$$P(X = 2) = P(X^{-1}(\{2\})) = P(\{(H,H)\}) = 1/4.$$

(b) In order to calculate the probability of having at least one head we have the following.

$$P(X \geq 1) = P(\{s \in S : X(s) \geq 1\}) = P(X^{-1}(\{1,2\})) = P(\{(H,T),(T,H),(H,H)\}) = 3/4.$$

9.7.1 Probability mass function (p.m.f.)

Definition 9.24 **Probability mass function**

Let (S,T,P) be a probability space and let X be a Z-valued discrete random variable on (S,T,P) $X : S \mapsto Z$. A probability mass function (p.m.f.) of the discrete r.v. X is the function f_X that associates a probability to each $s \in S$:

$$f_X : Z \mapsto [0,1],$$

where

$$f_X(z) = P(X = z), \forall z \in Z.$$

Note that $P(X = z)$ is equivalent to

$$f_X(z) = P(X = z) = P(\{s \in S : X(s) = z\}) = P(X^{-1}(\{z\})).$$

We used the notation $X = z$ instead of the more formal $\{X = z\}$. Then we would have written the previous set of equations as follows.

$$P(\{X = z\}) = P(\{s \in S : X(s) = z\}) = P(X^{-1}(\{z\})).$$

In other words the p.m.f. of r.v. X is a function that calculates $P(X = z)$ for each s in the domain of X with $X(s) = z$.

For a discrete random variable X that also takes discrete values $Z = \{x_1, \ldots, x_n\}$, we have that

$$P(X = x_i) = P(\{s \in S : X(s) = x_i\}) = f_X(x_i).$$

Example 9.7.3

For Example 9.7.2

$$f_X : \{0,1,2\} \mapsto [0,1],$$

where

$$f_X(z) = P(X = z) = \binom{2}{z}\frac{1}{4}.$$

Example 9.7.4

We have two dice. The sample space of rolling two dice is of cardinality 36, as there are so many pairs (d_1, d_2), where d_1 is one of the outcomes of one die, and likewise for d_2. Define random variable X as follows.

$$X : S \mapsto \mathbb{R} \quad \text{where } X(s) = X((d_1,d_2)) = d_1 + d_2, \quad s \in S.$$

Use the definitions to calculate $P(X = 4)$. The range of X is $R(S,X) = \{2,3,4,5,6,7,8,9,10,11,12\}$. Thus we can replace $\mathbb{R}$ with $Z = R(S,X)$ instead.

Proof. It is clear that $P(X = 4) = 1/12$. We properly define for $A \subseteq \mathbb{R}$ or $A \subseteq R(S,X)$, the set $X^{-1}(A) = \{s \in S : X(s) \in A\}$, that is the set of outcomes of S that result in getting mapped by X to a value in A. A will be set to $A = \{4\}$. This implies,

$$P(X \in A) = P(X^{-1}(A)).$$

The set $X^{-1}(A)$ must be measurable and thus it should belong to T i.e. $X^{-1}(A) \in T$. Say $A = \{4\}$. Then

$$X^{-1}(A) = \{(1,3),(2,2),(3,1)\}$$

Therefore

$$P(X \in A) = P(\{(1,3),(2,2),(3,1)\}) = 3/36 = 1/12.$$

We can also define similarly and analogously

$$P(X \in A) = \sum_{z \in A} f_X(z).$$

■

Theorem 9.5

Let (S,T,P) be a probability space and let X be a r.v. on Z. The events $\{X = z\}$ for $z \in Z$ are mutually exclusive and their union is S.

Proof. We need to show that events $\{X = z\}$ for $z \in Z$ are mutually exclusive with union in S. Consider $z_1, z_2 \in Z$ then for all s in S we have

$$\begin{aligned} s \in \{X = z_1\} \cap \{X = z_2\} &\Leftrightarrow s \in X^{-1}(\{z_1\}) \cap X^{-1}(\{z_2\}) \\ &\Leftrightarrow X(s) = z_1 \wedge X(s) = z_2 \\ &\Leftrightarrow z_1 = z_2. \end{aligned}$$

If $z_1 \neq z_2$ then $\{X = z_1\} \cap \{X = z_2\} = \emptyset$. So the events are mutually exclusive. Furthermore, $s \in S$ implies $s \in \{X = X(s)\}$ which gives $S = \cup_{z \in Z}\{X = z\}$. Moreover, if p.m.f of X is $f_X(z) = P(X = z)$,then $\sum_{z \in Z} f_X(z) = \sum_{z \in Z} P(X = z) = 1$. It follows from

$$\sum_{z \in Z} P(X = z) = P(\cup_{z \in Z}\{X = z\}) = P(S) = 1.$$

■

9.7.2 Composition of random variables

Definition 9.25 **Sum and product of a random variables**

Let X, Y be two random variables on the same probability space (S,T,P). Then $X+Y$, $X \cdot Y$ and $a \cdot X$, $a \in \mathbb{R}$ are also random variables and functions on S defined as follows.

$$\forall s \in S : (X+Y)(s) = X(s) + Y(s), \quad \forall s \in S : (X \cdot Y)(s) = X(s) \cdot Y(s), \quad \forall s \in S, a \in \mathbb{R} : (a \cdot X)(s) = aX(s).$$

Likewise for any polynomial function $f(x,y,\ldots,z)$ we define $f(X,Y,\ldots,Z)$ to be a function on S defined analogously.

$$\forall s \in S : f(X,Y,\ldots,Z)(s) = f(X(s),Y(s),\ldots,Z(s)).$$

9.7.3 Expectation

The term expectation is associated with a random variable. The terms mean and average can be used interchangeably with (mathematical) expectation, but the term mean usually refers to a distribution. The expectation of a random variable X would be denoted by $E(X)$ though it is more often to encounter $E[X]$, and errors might creep in. The term for mean is usually μ or $\bar{X}$.

Definition 9.26 **Expectation**

Let (S,T,P) be a probability space, with X a r.v. on Z and let $f_X(z) = P(X = z)$ be its p.m.f.. The expectation of X also known as the expected value of X, if it exists, is a real number $E(X)$ (or $E[X]$) defined as follows.

$$E(X) = E[X] = \sum_{s \in S} X(s)P(s).$$

Theorem 9.6

For a random variable X on Z and $f_X(z) = P(X = z)$, the following is given.

$$E(X) = E[X] = \sum_{s \in S} X(s)P(s) = \sum_{z \in Z} z f_X(z).$$

Proof. Note the following.

$$\begin{aligned} \sum_{s \in S} X(s)P(s) &= \sum_{z \in Z} \sum_{s, X(s)=z} X(s)P(s) \\ &= \sum_{z \in Z} zP(X = z) \\ &= \sum_{z \in Z} z f_X(z). \end{aligned}$$

■

Definition 9.27

Let (S, T, P) be a probability space, with X a r.v. on Z. Let X be a discrete random variable taking finite values $x_1, \ldots x_n$. Then the expectation of X, if it exists, is a real number $E(X)$ $(E[X])$ defined as follows.

$$E(X) = E[X] = \sum_{i=1}^{n} x_i P(X = x_i).$$

If X takes infinitely many values $x_1, x_2, \ldots$, then the expectation is defined as follows.

$$E(X) = E[X] = \sum_{i=1}^{\infty} x_i P(X = x_i).$$

Theorem 9.7

For a discrete and integer r.v. X on Z with p.m.f $f_X(x)$ the following applies.

$$E(X) = \sum_{i=0}^{\infty} P(X > i).$$

Proof. Consider $E(X) = \sum_j x_j P(X = x_j) = \sum_i iP(X = i)$. Note that if there is an integer x_j for which $P(X = x_j) \neq 0$, then let $x_j = k$ for some integer k. The term $x_j P(X = x_j)$ appears not only in the first sum but also in the second sum for $i = k = x_j$. For an i that appears in the second sum but not the first sum, the implication is that $P(X = i) = 0$ and thus why it is absent from the first sum, and in the second sum its contribution is $i \cdot P(X = i) = i \cdot 0$.

$$\begin{aligned} E(X) &= \sum_{j=1} x_j P(X = x_j) \\ &= \sum_{i=1} iP(X = i). \end{aligned}$$

We note the following

$$\begin{aligned}
E(X) &= \sum_{i=0}^{\infty} P(X > i) \\
&= P(X>0)+P(X>1)+P(X>2)+P(X>3)+\dots \\
&= P(X=1)+P(X=2)+P(X=3)+P(X=4)+\dots \\
&+ P(X=2)+P(X=3)+P(X=4)+\dots \\
&+ P(X=3)+P(X=4)+\dots \\
&+ P(X=4)+\dots \\
&= 1\cdot P(X=1)+2\cdot P(X=3)+3\cdot P(X=3)+4\cdot P(X=4)\dots \\
&= \sum_{i=1}^{\infty} iP(X=i).
\end{aligned}$$

■

Example 9.7.5

A coin is tossed twice. The sample space $S = \{HH, TH, HT, TT\}$. We count the number of heads and this becomes random variable X. Thus $R(S,X) = \{0,1,2\}$. The probabilities assigned by function $f(x_i) = P(X = x_i)$ are as follows.

$$P(X=0) = 1/4, P(X=1) = 1/2, P(X=2) = 1/4.$$

Then the expection of r.v. X becomes

$$\mu = E(X) = 0\cdot P(x=0)+1\cdot P(X=1)+2\cdot P(X=2) = 0\cdot 1/4+1\cdot 1/2+2\cdot 1/4 = 1$$

(We thus expect half of the tosses to be H.)

Definition 9.28 **Expectation of $g(X)$**

Let (S,T,P) be a probability space, with X a r.v. on Z and let $f_X(z) = P(X = z)$ be its p.m.f.. The expectation of $g(X)$, if it exists, is a real number $E(g(X))$ defined as follows.

$$E(g(X)) = \sum_{s\in S} g(X(s))P(s).$$

Theorem 9.8

For a function $g(X)$ as defined above,

$$E(g(X)) = \sum_{s\in S} g(X(s))P(s) = \sum_{x} g(x)f_X(x).$$

Proof.

$$\begin{aligned}
E(g(X)) &= \sum_{s\in S} g(X(s))P(s) \\
&= \sum_{s\in S} g(X(s))P(\{s\}) \\
&= \sum_{x}\sum_{s\in S, X(s)=x} g(X(s))P(\{s\}) \\
&= \sum_{x}\sum_{s\in S, X(s)=x} g(x)P(\{s\}) \\
&= \sum_{x} g(x)\sum_{s\in S, X(s)=x} P(\{s\}) \\
&= \sum_{x} g(x)P(X=x) \\
&= \sum_{x} g(x)f_X(x).
\end{aligned}$$

■

9.7.4 Joint probability

Definition 9.29

Let (S,T,P) be a probability space, and two random variables X and Y. Let $f_X(x)$ and $f_Y(y)$ be the p.m.f of X and Y respectively. The $f_{X,Y}(x,y)$ is the joint probability of X and Y. It is the case,

$$f_X(x) = \sum_y f_{X,Y}(x,y), \text{ and } f_Y(y) = \sum_x f_{X,Y}(x,y).$$

The joint probability $f_{X,Y}(x,y)$ is the probability $P(\{s : X(s)=x\}\cap\{s : Y(s)=y\})$ that is usually simplified in the following form $P(\{X=x\}\cap\{Y=y\})$ or $P(\{X=x\}\wedge\{Y=y\})$ or $P(\{X=x\},\{Y=y\})$. We summarize all these forms as follows.

$$\begin{aligned}
f_{X,Y}(x,y) &= P(\{s : X(s)=x\}\cap\{s : Y(s)=y\}) \\
&= P(\{X=x\}\cap\{Y=y\}) \\
&= P(X=x\wedge Y=y) \\
&= P(X=x, Y=y).
\end{aligned}$$

Example 9.7.6

(a) Consider the following experiment. A urn contains three balls colored R, G, B. We draw two balls from the urn (without replacement). The outcome of the first drawing is represented by r.v. X and of the second by Y. Calculate $f_{X,Y}(x,y)$ and $f_X(x)$.
(b) Repeat (a) when we draw two balls from the run with replacement: after we draw the first ball and record the outcome (color) it is placed back into the urn.

Proof.
(a) $f_{X,Y}(x,y)$ is displayed on the table.

$f_{X,Y}(x,y)$	R	G	B
R	0	1/6	1/6
G	1/6	0	1/6
B	1/6	1/6	0

It is clear that $f_X(R) = f_X(G) = f_X(B) = 1/3$. Moreover,

$$\begin{aligned} f_X(R) &= \sum_{y\in S} f_{X,Y}(x,y) = 0+1/6+1/6 = 1/3, \\ f_X(G) &= \sum_{y\in S} f_{X,Y}(x,y) = 1/6+0+1/6 = 1/3, \\ f_X(B) &= \sum_{y\in S} f_{X,Y}(x,y) = 1/6+1/6+0 = 1/3. \end{aligned}$$

(b) $f_{X,Y}(x,y)$ is displayed on the table.

$f_{X,Y}(x,y)$	R	G	B
R	1/9	1/9	1/9
G	1/9	1/9	1/9
B	1/9	1/9	1/9

■

Theorem 9.9

Let (S,T,P) be a probability space, and two random variables X and Y. Let $f_X(x)$ and $f_Y(y)$ be the p.m.f of X and Y respectively. Then, $f_{X,Y}(x,y)$ is the joint probability of X and Y. The expectation of $g(X,Y)$, if it exists, is a real number $E(g(X,Y))$ defined as follows.

$$E(g(X,Y)) = \sum_{x,y} g(x,y)P(X=x,Y=y).$$

Proof. Let $W = g(X,Y)$. For $f_W(w)$ the following is applies.

$$\begin{aligned} f_W(w) &= P(W=w) \\ &= \sum_{x,y:g(x,y)=w} P(g(X,Y)=w) \\ &= \sum_{x,y:g(x,y)=w} P(X=x,Y=y) \\ &= \sum_{x,y:g(x,y)=w} f_{X,Y}(X,Y). \end{aligned}$$

Then we have the following.

$$\begin{aligned} E(W) &= \sum_w w f_W(w) \\ &= \sum_w w P(W=w) \\ &= \sum_w w P(g(X,Y)=w) \\ &= \sum_w w \sum_{x,y,w:w=g(x,y)} P(X=x,Y=y) \\ &= \sum_{x,y} \sum_{w:w=g(x,y)} w P(X=x,Y=y) \\ &= \sum_{x,y} g(x,y)P(X=x,Y=y). \end{aligned}$$

■

9.7.5 Indepdendent random variables

Definition 9.30 **Independent discrete r.v.**

Let X and Y be two random variables taking values $x_1,\ldots,x_n$ and $y_1,\ldots,y_m$ respectively. We say X and Y are independent (random variables), if for every i and j, we have the following

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j).$$

In other words X and Y are independent if the events $\{X=x_i\}$ and $\{Y=y_j\}$ are independent.

Theorem 9.10

If X and Y are two discrete random variables for a probability space (S,T,P) then if X and Y are independent we have the following.

$$E(XY)=E(X)\cdot E(Y),$$

The result extends for continuous r.v. X and Y.

Proof. Let $X=\cup\{x_i\}$, and $Y=\cup\{y_j\}$. Using prior notation $P_X(X=x_i)=f_X(x_i)$ and $P_Y(Y=y_j)=f_Y(x_j)$. If X and Y are independent, we have $P(X=x_i\cap Y=y_j)=P(X=x_i)\cdot P(Y=y_j)$.

$$\begin{aligned}
E(XY) &= \sum_{i,j} x_i y_j P(X=x_i\cap Y=y_j)\\
&= \sum_{i,j} x_i y_j P(X=x_i)\cdot P(\cap Y=y_j)\\
&= \sum_{i} x_i P(X=x_i)\cdot + \sum_{j} y_j P(Y=y_j)\\
&= E(X)\cdot E(Y).
\end{aligned}$$

■

Example 9.7.7

Trom a prior example of dice tossing. Random variable Y is defined as follows

$$Y:S\to\mathbb{R} \quad \textbf{where } Y((a,b))=\min(a,b).$$

Thus for an element $s=(a,b)$ of S, random variable Y assigns a value to this element that is equal to the minimum of the two values of the two faces of the two dice. The range of Y is $R(S,Y)=\{1,2,3,4,5,6\}$. We can compute

$$P(Y=1)=11/36, P(Y=2)=9/36, P(Y=3)=7/36, P(Y=4)=5/36, P(Y=5)=3/36, P(Y=6)=1/36$$

For example, $P(Y=1)$ is derived from the fact that the tosses $(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)$ have 1 as the minimum value. Furthermore the following outcomes $(4,4),(4,5),(4,6),(5,4),(6,4)$ have all minimum value equal to 4, and so on. Thus

$$\mu=E(Y)=1\cdot 11/36+2\cdot 9/36+3\cdot 7/36+4\cdot 5/36+5\cdot 3/36+6\cdot 1/36=\frac{11+18+21+20+15+6}{36}=2.527.$$

9.7.6 Variance

The variance of a random variable is the expected value of the squared deviation from the mean of the random variable.

The variance of X is denoted $Var(X)$ or $Var(X)$ or $\sigma^2(X)$ or just σ^2 if the reference to X is obvious. Then $\sqrt{\sigma^2(X)} = \sigma(X)$ or σ is the standard deviation of X.

Definition 9.31

Let (S,T,P) be a probability space, with X a r.v. on Z. Let X be a discrete random variable taking finite values $x_1,\ldots x_n$. Then the variance of X, if it exists, is a real number $Var(X)$ $(Var[X])$ defined as follows.

$$Var(X) = \sigma^2(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.$$

Note that

$$E(X^2) = E[X^2] = \sum_{i=1}^{n} x_i^2 P(X = x_i),$$

or

$$E(X^2) = E[X^2] = \sum_{i=1}^{\infty} x_i^2 P(X = x_i),$$

as needed. The standard deviation of X is defined as $\sqrt{var(X)} = \sigma(X)$.

The variance exists if the second momdent $E(X^2)$ exists. The latter implies that the first moment $E(X)$ also exists.

Corollary 9.2

For $var(X)$ of a random variable X we have

$$var(X) = E(X^2) - (E(X))^2.$$

Proof.

$$var(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + (E(X))^2) = E(X^2) - 2(E(X))^2 + (E(X))^2) = E(X^2) - (E(X))^2.$$

Another way to prove this for a discrete r.v. X is as follows.

$$\begin{aligned} E((X - E(X))^2) &= \sum_{x_i \in X} (x_i^2 + E(X) * E(X) - 2x_i E(X)) f(x_i) \\ &= \sum_{x_i \in X} (x_i^2 f(x_i)) + \sum_{x_i \in X} (E(X) \cdot E(X)) f(x_i) + \sum_{x_i \in X} (-2x_i E(X)) f(x_i) \\ &= E(X^2) + E(X) \cdot E(X) \cdot \sum_{x_i \in X} f(x_i) - 2E(X) \cdot \sum_{x_i \in X} x_i f(x_i) \\ &= E(X^2) + E(X) \cdot E(X) - 2E(X) \cdot E(X) \\ &= E(X^2) - E(X) \cdot E(X). \end{aligned}$$

For the continuous x a similar proof can be obtained. ■

Theorem 9.11

Let X be a discrete random variable with $E(X) = \mu$. Then

$$Var(X) = E((X - \mu)^2) = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i).$$

Note that $E(X)=\mu$, $\sum_i P(X=x_i)=1$, and thus

$$E((X-\mu)^2)=E(X^2)-2\mu E(X)+\mu^2=E(X^2)-E(X)^2.$$

Theorem 9.12

Let X and Y be two discrete random variables. The following apply.

$$E(X+Y)=E(X)+E(Y),$$

$$Var(X+Y)=Var(X)+Var(Y), \quad \text{for independent} \quad X,Y.$$

Moreover,

$$E(aX)=aE(X), \quad Var(aX)=a^2Var(X),$$

for a real constant $a\neq 0$.

Proof. Let us assume that X,Y are discrete random variables. Then X takes values $x_1,x_2,\ldots$ and Y takes values $y_1,y_2,\ldots$, and therefore $X+Y$ takes values $x_i+y_j,\ldots,1\leq i,j\ldots$, or equivalently $X+Y$ takes values $z_k=x_i+y_j,\ldots,1\leq i,j\ldots$.

$$\begin{aligned}
E(X+Y) &= \sum_k z_k P(X+Y=z_k)\\
&= \sum_i\sum_j (x_i+y_j)P(X=x_i,Y=y_j)\\
&= \sum_i x_i\sum_j y_j P(X=x_i,Y=y_j)+\sum_j y_j\sum_i x_i P(X=x_i,Y=y_j)\\
&= \sum_i x_i P(X=x_i)+\sum_j y_j P(Y=y_j)\\
&= E(X)+E(Y).
\end{aligned}$$

For the variance we first calculate

$$E((X+Y)^2)=E(X^2+Y^2+2XY)=E(X^2)+E(Y^2)+2E(XY).$$

We also consider the fact that X,Y are independent.

$$\begin{aligned}
E(XY) &= \sum_k z_k P(XY=z_k)\\
&= \sum_i\sum_j (x_i * y_j)P(X=x_i,Y=y_j)\\
&= \sum_i x_i\sum_j y_j P(X=x_i)P(Y=y_j)\\
&= \sum_i x_i P(X=x_i)\sum_j y_j P(Y=y_j)\\
&= E(X)\cdot E(Y).
\end{aligned}$$

Then

$$\begin{aligned}
Var(X+Y) &= E((X+Y)^2)-(E(X+Y))^2\\
&= E(X^2)+E(Y^2)+2E(XY)-(E(X)+E(Y))^2\\
&= E(X^2)+E(Y^2)+2E(X)E(Y)-E^2(X)+E^2(Y)-2E(X)E(Y)\\
&= Var(X)+Var(Y).
\end{aligned}$$

Furthermore,

$$E(aX) = \sum_i ax_i P(aX = ax_i) = \sum_i ax_i P(X = x_i) = aE(X).$$

$$Var(aX) = E((aX)^2) - E(aX)^2 = a^2E(X^2) - a^2E(X)^2 = a^2Var(X).$$

■

It becomes clear then that $Var(aX+b) = a^2Var(X)$.

Corollary 9.3

For a r.v. X and any $a, b \in \mathbb{R}$, we have the following.

$$var(aX+b) = a^2var(X).$$

9.7.7 Cumulative distribution function

We can also define a cumulative distribution function for a discrete r.v. X. Then

$$F_X(x) = \sum_{t=X(s)\leq x, s\in S} f_X(t).$$

Definition 9.32 **Distribution of a random variable**

For a random variable X let $f_X(x)$ be it p.m.f. We can define $F_X(x)$ its cumulative distribution function (c.d.f.) as follows.

$$F_X(x) = P(X \leq x) = \sum_{t\leq x} f_X(t).$$

Definition 9.33

Let X be a random variable taking values $x_1, \ldots, x_n$. Let $x_1 < x_2 < \ldots < x_n$. The cumulative distribution function (c.d.f.) of X is given as follows.

$$F_X(x_i) = P(X \leq x_i) = \sum_{m=1}^{i} P(X = x_i).$$

Moreover,

$$P(X = x_i) = F_X(x_i) - F_X(x_{i-1}).$$

9.7.8 Probability generating function

Definition 9.34

Let X be a random variable taking values $x_1, \ldots, x_n$. The probability generating function of X is the power series defined as follow.

$$g_X(x) = \sum_i P(X = x_i)x^i.$$

Theorem 9.13

Let X be a random variable taking values $x_1, \ldots, x_n$. Let $g_X(x)$ be the probability generating function of X. The following apply.

$$(a) g_X(1) = \sum_i P(X = x_i) = 1,$$

$$(b) E(X) = [(g_X(x))']_{x=1},$$

and

$$(c) Var(X) = [(g_X(x))^{''}]_{x=1} + E(X) - E(X)^2.$$

Proof.
(a) Direct derivation.
(b) By taking a first derivative, we obtain

$$(g_X(x))' = \sum_i P(X = x_i) \cdot i \cdot x^{i-1}$$

Setting $x = 1$ we then obtain the following.

$$(g_X(1))' = \sum_i P(X = x_i) \cdot i = E(X).$$

(c) By differentiating twice, we obtain the following.

$$(g_X(x))^{''} = \sum_i P(X = x_i) \cdot i \cdot (i-1) \cdot x^{i-2}$$

Setting $x = 1$ we then obtain the following.

$$(g_X(1))^{''} = \sum_i P(X = x_i) \cdot (i^2 - i) == E(X^2) - E(X)$$

Adding $E(X) - E(X)^2$ we obtain the following.

$$(g_X(1))^{''} + E(X) - E(X)^2 == E(X^2) - E(X) + E(X) - E(X)^2 = Var(X).$$

■

9.8 Continuous random variables

Definition 9.35 **Continuous r.v.**

A random variable is continuous if its domain is uncountably infinite.

9.8.1 Cumulative distribution function

Definition 9.36 **Distribution of a random variable**

The cumulative distribution function of a continuous r.v. X is described as the distribution of a random variable and is denoted by $F_X(x)$ (sometimes, just $F(x)$).

$$F_X(x) = F(x) = P(X \leq x) = P(X^{-1}((-\infty, x])).$$

We note that $0 \leq F_X(x) \leq 1$.

Given X the cumulative distribution function F can be defined as $F_X(x) = P(X \leq x)$, for $-\infty < x < \infty$. One can also define $F_X(x) = P(A)$, where $A = (-\infty, x]$, where $x \in \mathbb{R}$. Function $F_X(x)$ or simply $F(x)$ when it implicitly infers to X, is a monotonically increasing continuous from left function with

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$$

Furthermore, if there is $f(t)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

then $f(t)$ is the probability density function of X. Function $f(t)$ to exist it means that $F(x)$ is differentiable. In the remainder, as needed, it will be assumed that $F(x)$ is differentiable. An interpretation of the density function is

$$f(x)dx \approx P(x < X \leq x + dx).$$

Theorem 9.14

Let (S, T, P) be a probability space. The (cumulative) distribution function $F_X(x)$ of r.v. X has the following properties.

1. *F is nondecreasing.*
2. *$\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$,*
3. *F is right continuous $\lim_{y \to x^+} F(y) = F(x)$,*
4. *$P(X < x) = \lim_{y \to x^-} F(y)$,*
5. *$P(X = x) = F(x) - P(X < x)$.*

F is nondecreasing since for $x < y$,

$$F_X(y) - F_X(x) = P(X \leq y) - P(X \leq x) = P(x < X \leq y) \geq 0.$$

The distribution of X is called continuous if there is a function $f_X(x)$ (called the probability density function) such that

$$F_X(B) = \int_B f_X(x)dx,$$

for every interval B (or in mathematical analysis terms, for every Borel set B) and in that case F_X is continuous.

9.8.2 Probability density function (p.d.f.)

Definition 9.37 **Probability density function**

When the distribution function $F_X(x) = P(X \leq x)$ has the form

$$P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt,$$

we refer to $f_X(x)$ as the probability density function of X and sometimes we denote it $f(x)$ implicitly referring to r.v. X. Moreover,

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

If $F(x)$ is differentiable, then $f(x)$ exists. It is possible however that $F(x)$ is not differentiable.

Moreover

$$P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(t)dt.$$

9.8.3 Expectation of a continuous r.v.

Definition 9.38 **Expectation of a continuous r.v.**

Let X be a continuous and differentiable random variable on probability space (S,T,P), with probability density function $f_X(x)$. Then the expectation or (expected value) of r.v. X is defined as follows.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx.$$

If $E(X) = -\infty$ or $E(X) = \infty$ we say the expectation $E(X)$ does not exist.

A function f is absolutely continuous say on a closed interval $[a,b]$ if it has a derivative f' almost everywhere, the derivative is Lebesgue integrable and thus $f(x) = f(a)\int_a^x f'(z)dz$ for all $x \in [a,b]$.

Definition 9.39 **Expectation of g(X)**

For an absolutely continuous random variable X, with probability density function $f_X(x)$, let $g : \mathbb{R} \mapsto \mathbb{R}$. Then the expectation of $g(X)$ exists and

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx.$$

Definition 9.40 **Expectation of h(X,Y)**

For jointly absolutely continuous random variables X,Y, with joint probability density function $f_{X,Y}(x,y)$, let $h : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$. Then the expectation of $h(X,Y)$ exists and

$$E(h(X,Y)) = \int_{-\infty}^{\infty} h(X,Y) f_{X,Y}(x,y)dx.$$

9.8.4 Independent random variables: continuous case

Definition 9.41 **Independent r.v.: continuous case**

Let X and Y be two random variables with $F_X(x)$ and $F_Y(y)$ distribution functions. Let $f_{X,Y}(x,y)$ be their joint distribution function. Then X and Y are independent r.v. if, for all x,y,

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

9.8.5 Variance of a continuous r.v.

Definition 9.42 **Variance of a continuous r.v.**

For a continuous random variable X, its variance $Var(X)$ is defined as follows.

$$\sigma_X^2 = Var(X) = E((X-\mu_X)^2) = E((X-E(X))^2) = E(X^2) - E(X),$$

where

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx.$$

σ_X is known as the standard deviation of X, and μ_X is the expectation (mean) of X.

Theorem 9.15

Let X be a continuous r.v. with $E(X)$ and $Var(X)$. The following apply.

1. $Var(X) \geq 0$.
2. $Var(aX+b) = a^2 Var(X)$, *for any real* $a, b \in \mathbb{R}$.
3. $Var(X) = E(X^2) - E(X)^2$.
4. $Var(X) \leq E(X^2)$.

Proof.
(1) Follows from the definition $Var(X) = E((X-\mu_X)^2) \geq 0$.
(2) By linearity of expectation $E(aX+b) = aE(X)+b$. Therefore

$$\begin{aligned} Var(aX+b) &= E((aX+b) - E(aX+b))^2) \\ &= E((aX+b) - aE(X)+b)^2) \\ &= E((aX - aE(X))^2) \\ &= a^2 E((X-E(X))^2) \\ &= a^2 Var(X). \end{aligned}$$

(3) Using again the linearity of expectation the following is derived.

$$\begin{aligned} Var(X) &= E((X-E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

(4) By the previous part

$$Var(X) = E(X^2) - E(X)^2 \leq E(X^2),$$

since the last term $E(X)^2$ is positive. ■

Definition 9.43 **Covariance**

The covariance of two random variable X, Y is given by the following expression.

$$Cov(X,Y) = E((X-E(X)) \cdot (Y-E(Y))).$$

Theorem 9.16

Let X,Y be two random variables. Then

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

Proof.

$$Cov(X,Y) = E((X - E(X)) \cdot (Y - E(Y))) = E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y).$$

■

Theorem 9.17

Let X,Y be two random variables. Then

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y).$$

The result is generalized as follows.

$$Var(\sum_i X_i) = \sum_i Var(X_i) + 2\sum_{i<j} Cov(X_i,X_j).$$

Proof.

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)2 - 2E(X)E(Y) \\ &= var(X) + var(Y) + 2(E(XY) - E(X)E(Y) \\ &= var(X) + var(Y) + 2cov(X,Y). \end{aligned}$$

■

If X and Y are indepdendent $E(XY) = E(X)E(Y)$ and thus $cov(X,Y) = 0$. We the obtain the following.

Corollary 9.4

For r.v. X, Y that are independent we have the following.

$$var(X+Y) = var(X) + var(Y).$$

Definition 9.44 **Correlation**

The correlation of two random variable X,Y is given by the following expression.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

9.8.6 Median and other quartiles

Definition 9.45

For a random variable X, the median m, lower quartile l, and upper quartile u are defined as follows.

$$F_X(m) = 1/2, \quad F_X(q) = 1/4, \quad F_X(u) = 3/4.$$

The n-th percentile of X is defined as the valye x_n such that

$$F_X(x_n) = n/100.$$

9.8.7 Miscellanea

Fact 9.1

From prior discussion $n! \leq n^n$ and $n! \approx (n/e)^n$. Moreover

$$(n/e)^n \leq n! \leq en(n/e)^n$$

Also for $\binom{n}{k}$ we have $\binom{n}{k} \leq n^k$ or

$$(n/k)^k \leq \binom{n}{k} \leq (ne/k)^k$$

For all $k \leq n$ we have

$$\binom{n}{k} \leq 2^n.$$

Furthermore for $\binom{2n}{n}$ we have

$$\frac{2^{2n}}{2\sqrt{n}} \leq \binom{2n}{n} \leq \frac{2^{2m}}{\sqrt{2m}}.$$

Fact 9.2

For all real x $e^x \geq 1+x$. Moreover $(1-x)^n$ for $x > 0$ and small $(1-x)^n \leq \exp -np$. Also $1-x \geq e^{-2x}$ if $x \leq 1/2$. For all $x > 1$ we have $1 - 1/x < 1/e$.

9.9 Standard distributions and random variables

9.9.1 Uniform

Definition 9.46

Uniform U(a,b) random variable. Let a,b be real numbers with $a<b$. A uniform random variable $X\sim U(a,b)$ on the interval $[a,b]$ has p.d.f as follows.

$$f_X(x)=1/(b-a),\quad a\le x\le b,\quad f_X(x)=0\quad \text{otherwise.}$$

Moreover,

$$F_X(x)=0 \text{ if } x<a,\quad F_X(x)=1 \text{ if } x>b,\quad \textit{and}\quad F_X(x)=\frac{x-a}{b-a} \text{ if } a\le x\le b.$$

Furthermore

$$E(X)=(a+b)/2,$$

and

$$Var(X)=(b-a)^2/12.$$

9.9.2 Bernoulli distribution

Definition 9.47 **Bernoulli Trials**

A Bernoulli trial is an experiment that has two outcomes. One outcome is called a success and the other a failure. Let p be the probability of success and $q=1-p$ be the probability of failure. We denote such Bernoulli trial with $b(p)$.

Several times 1 indicates success and 0 indicates a failure in the context of a Bernoulli trial.

If however $1/2$ indicates success and $-1/2$ a failure we call the experiment a Rademacher trial.

Corollary 9.5

A r.v. X follows a Bernoulli distribution that is $X\sim b(p)$ and it has the following properties.

$$E(X)=p,\qquad var(X)=pq=p(1-p),$$

where $q=1-p$.

Proof. For $E(X)$ we have the following.

$$E(X)=p\cdot 1+(1-p)\cdot 0=p$$

Furthermore for $E(X)$ we have the following.

$$E(X^2)=p\cdot 1^2+(1-p)\cdot 0=p$$

Then for $var(X)$ we have the following.

$$var(X)=E(X^2)-(E(X))^2=p-p^2=p(1-p)=pq.$$

■

9.9.3 Binomial distribution

Definition 9.48 **Binomial process or trial**

A Binomial process or trial is the independent repetition of identical Bernoulli trials. Independent means the outcome of a Bernoulli trial (or experiment) is not dependent of previous outcomes. We denote such a Binomial process with $B(n,p)$, where n indicates the number of Bernoulli trials $b(p)$, and p the property of the Bernoulli trial (probability of success).

Let $B(n,p)$ denote a binomial process of n independent Bernoulli trials each one $b(p)$.

Theorem 9.18

The probability of k successes in a binomial experiment $B(n,p)$ is denoted by $B(n,p;k)$ and given by

$$B(n,p;k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Corollary 9.6

The probability of one or more success is $1 - B(n,p;0) = 1-(1-p)^n = 1-q^n$, where $q = 1-p$ is the probability of failure.

Corollary 9.7

We toss a (fair) coin 8 times. Thus $p = q = 1/2$. The probability of no heads is $1/2^8 = 1/256$. The probability of at least one head is is $1-(1-1/2)^8 = 1-1/256$.

Theorem 9.19 **Binomial Distribution $B(n,p)$**

For the binomial distribution, we have that $P(X=k) = B(n,p;k) = \binom{n}{k} p^k q^{n-k}$, where $q = 1-p$. Then

$$\mu = E[X] = \sum_k k P(X=k) = np.$$

$$Var(X) = E[(X-\mu)^2] = npq.$$

$$\sigma(X) = \sigma_X = \sqrt{npq}.$$

Theorem 9.20

Let X be a random variable that follows a binomial distribution with parameters n and p. Sometimes we use $q = 1-p$. We write that $X \sim B(n,p)$. The probability of k successes in n trials is B(n,p;k). The expected value of $E(X)$ and its variance $var(X) = \sigma^2(X)$ are as follows.

$$E(X) = np, \qquad var(X) = \sigma^2(X) = npq.$$

Proof. We will use the binomial theorem that states

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j},$$

and therefore also

$$(a+b)^{n-1} = \sum_{j=0}^{n-1} \binom{n-1}{j} a^j b^{n-1-j},$$

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \cdot Pr(X=k) = \sum_{k=0}^{n} k \cdot B(n,p;k) \\
&= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= n \cdot p \cdot \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= n \cdot p \cdot \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} (1-p)^{n-1-(k-1)} \\
&= np \cdot \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j} \\
&= np \cdot (1-p+p)^{n-1} = np.
\end{aligned}$$

The binomial theorem was used in the last step. We then calculate $E(X^2)$. When we split the major sum, the latter second term is $E(X)$ above.

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \cdot B(n,p;k) \\
&= \sum_{k=0}^{n} (k^2 - k + k) \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} (k^2 - k) \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= p^2 \cdot n \cdot (n-1) \cdot \sum_{k=2}^{n} k(k-1) \cdot \frac{(n-2)!}{k(k-1) \cdot (k-2)! \cdot (n-k)!} p^{k-2} (1-p)^{n-k} + np \\
&= p^2 \cdot n \cdot (n-1) \cdot \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)! \cdot (n-k)!} p^{k-2} (1-p)^{n-2-(k-2)} + np \\
&= p^2 \cdot n \cdot (n-1) \cdot \sum_{j=0}^{n-2} \frac{(n-2)!}{j! \cdot (n-j)!} p^j (1-p)^{n-2-j} + np \\
&= n(n-1) p^2 \cdot (p+1-p)^{n-2} + np \\
&= n^2 p^2 + np(1-p).
\end{aligned}$$

Finally, we combine the two results for $E(X^2)$ and $E(X)$ in the definition of variance to obtain.

$$var(X) = E(X^2) - E^2(X) = n^2 p^2 + np(1-p) - (np)^2 = np(1-p) = npq.$$

■

9.9.4 Normal

Definition 9.49

Normal $N(\mu, \sigma^2)$ random variable. A normal random variable $X \sim N(\mu, \sigma^2)$ has p.d.f as follows.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-(x-\mu)^2 / 2\sigma^2\right)$$

Furthermore

$$E(X) = \mu$$

and

$$Var(X) = \sigma^2.$$

If $X \sim N(\mu, \sigma^2)$ the $Y = (X - \mu)/\sigma \sim N(0,1)$.

9.9.5 Geometric distribution

Definition 9.50

A geometric distribution can be described as the number of Bernoullis trials to get the first success, or equivalently the number of failures before the first success. If random variable X describes the number of Bernoulli trials and variable Y the number of failure to the first success, then $X = Y + 1$. We then say $X \sim g(p)$, where p is the probability of success in a Bernoulli trial.

Corollary 9.8

A r.v. X follows a geometric distribution that is $X \sim g(p)$ and it has the following properties.

$$E(X) = 1/p, \qquad var(X) = (1-p)/p^2.$$

where $q = 1 - p$. Note that $P(X = k) = (1-p)^{k-1} \cdot p$.

Proof. Let $X \sim g(p)$. One can show $E(X) = 1/p$ and $var(X) = (1-p)/p^2$. A straightforward method to derive the expectation is to argue that with probability p we are done in one trial, otherwise with probability $1-p$ we will will need one trial (the just wasted trial) plus $E(X)$ for the remaining trials. Then $E(X) = p + (1-p)(1+E(X))$ gives $E(X) = 1/p$. Otherwise, Start with $P(X = k) = (1-p)^{k-1}p$. Then

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} kP(X=k) \\ &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\ &= p/(1-p) \cdot \sum_{k=1}^{\infty} k(1-p)^k \\ &= p/(1-p) \cdot \sum_{k=0}^{\infty} k(1-p)^k \\ &= \frac{p}{1-p} \cdot \frac{1-p}{p^2} = \frac{1}{p}. \end{aligned}$$

The expected number of failures is one less $1/p - 1 = (1-p)/p$. Thus $E(Y) = (1-p)/p$ as well. For $var(X) = E(X^2) - (E(X))^2$ we only need compute the first term $E(X^2)$. We have from sequences $\sum_{i=0}^{\infty} x^i = 1/(1-x)$ for $x < 1$. Then $\sum_{i=1}^{\infty} ix^{i-1} = 1/(1-x)^2$ for $x < 1$, that was used before for $E(X)$. Then $\sum_{i=2}^{\infty} i(i-1)x^{i-2} = 2/(1-x)^3$

for $x < 1$.

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{\infty} k^2 P(X=k) \\ &= \sum_{k=1}^{\infty} (k^2 - k + k)(1-p)^{k-1} p \\ &= \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1} p + \sum_{k=1}^{\infty} k(1-p)^{k-1} p \\ &= p(1-p) \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + E(X) \\ &= p(1-p)(2/(1-(1-p))^3) + 1/p \\ &= 1/p + 2(1-p)/p^2 \end{aligned}$$

Then

$$var(X) = E(X^2) - (E(X))^2 = 1/p + 2(1-p)/p^2 - (1/p)^2 = (1-p)/p^2.$$

■

9.10 Generating functions

9.10.1 Probability generating function

Definition 9.51

Let X be a discrete random variable. Then the probability generating function (p.g.f.) is denoted by

$$p_X(t) = E(t^X),$$

for $t \in \mathbb{R}$.

Theorem 9.21

Let X be a discrete random variable. Let $X : S \mapsto Z$, and Z is a subset of $\mathbb{N}$. For $p_X(t)$ let there exists ρ such that $p_X(\rho) < \infty$. Then the following apply.

$$\begin{aligned} p_X(0) &= P(X=0) \\ p'_X(0) &= 1! \cdot P(X=1) \\ p'''_X(0) &= 2! \cdot P(X=2) \\ &\dots \\ p_X^{(m)}(0) &= m! \cdot P(X=m). \end{aligned}$$

Proof.

$$p_X(t) = E(t^X) = \sum_{k=0}^{\infty} t^k P(X=k) = t^0 P(X=0) + t^1 P(X=1) + t^2 P(X=2) + \ldots + t^m P(X=m) + \ldots$$

Set $t = 0$. Then $p_X(0) = P(X = 0$ since all term but the first vanish. Take m time the derivative of both sides. The following applies.

$$p_X^{(m)}(t) = m! P(X=m) + O(1)t(P(X=m+1) + \ldots).$$

Seeting $t = 0$ the second summand vanishes and the general result is obtained. ■

9.10.2 Moments

Definition 9.52

Let X be a discrete random variable and m a non-negative natural number. The m-th moment of X is $E(X^m)$; it exists provided this given expectation exists.

9.10.3 Moment generating function

Let $\exp(Z)$ denote e^Z.

Definition 9.53

The moment generating function (m.g.f.) $m_X(t)$ of a random variable X, with cumulative distribution function F_X is defined as follows, provided that the expectation below is defined for t in some neighborhood of 0 in other words there exists an $h > 0$ such that for all t in $[-h, h]$, $E(\exp(tX))$ exists.

$$m_X(t) = E(e^{tX}),\ t \in \mathbb{R}.$$

For a continuous r.v. we have than

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx.$$

Theorem 9.22

Let X be any random variable. Then $m_X(t) = p_X(e^t)$.

Note that using Taylor's expansion

$$\exp(tX) = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \ldots + \frac{t^nX^n}{n!} + \ldots.$$

Taking expectations of both sides we derive the following.

$$E(\exp(tX)) = 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \frac{t^3E(X^3)}{3!} + \ldots + \frac{t^nE(X^n)}{n!} + \ldots.$$

Differentiating n times $m_X(t)$ with respect to t and setting $t = 0$ we obtain the n-th moment $E(X^n)$.

Theorem 9.23

Let X be any random variable. Let for some $\rho > 0$ we have $m_X(t) < \infty$ for $-\rho < t < rho$. Then the following apply.

$$\begin{aligned} m_X(0) &= 1 \\ m'_X(0) &= E(X) \\ m'''_X(0) &= E(X^2) \\ &\ldots \\ m_X^{(m)}(0) &= E(X^m). \end{aligned}$$

Proof.
(a) $m_X(0) = p_X(e^0) = E((e^0)^X) = E(1) = 1$.
(b) Take m time the derivative of $m_X(t) = E(e^{tX})$. The first time

$$m'_X(t) = E(Xe^{tX}) \Leftrightarrow m'_X(0) = E(Xe^{0 \cdot X}) = E(X).$$

The second time

$$m_X^{(1)}(t) = E(Xe^{tX}) \Leftrightarrow m_X^{(2)}(t) = E(X^2e^{t \cdot X}) \Leftrightarrow m_X^{(2)}(0) = E(X^2e^{0 \cdot X}) = E(X^2).$$

By induction, the general term is derived. ■

Corollary 9.9

If X, Y are independent,

$$m_{X+Y}(t) = m_X(t)m_Y(t), \quad p_{X+Y}(t) = p_X(t)p_Y(t),$$

Theorem 9.24

Let X_i be independent random variables, with m.g.f $E_{X_i}(t)$ respectively, $i = 1, \ldots, n$. Let $S_n = \sum_{i=1}^n X_i$. Then the following applies for the m.g.f. of random variable Y.

$$m_Y(t) = M_{X_1}(t) \ldots M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Furthermore, if X_i are identically distributed to random variable X, then

$$m_Y(t) = (M_X(t))^n.$$

9.10.4 Some examples

Example 9.10.1
The Uniform distribution $U(a,b)$ has

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Example 9.10.2
The binomial distribution $B(n,p)$ has

$$M_X(t) = (1-p+pe^t)^n.$$

Example 9.10.3
The normal distribution $N(\mu, \sigma^2)$ has

$$M_X(t) = \exp(\mu t + 0.5\sigma^2 t^2).$$

Example 9.10.4
The Poisson distribution $Poisson(\lambda)$ has

$$M_X(t) = \exp(\lambda(e^t - 1)).$$

Example 9.10.5
The geometric distribution $g(p)$ with parameter p, has

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t} = \frac{pe^t}{1-e^t+pe^t}.$$

Proof. Let $X \sim g(p)$. One can show $E(X) = 1/p$ and $var(X) = (1-p)/p^2$. For the moment generative function we observe the following

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk}(1-p)^{k-1}p \\ &= pe^t \sum_{k=1}^{\infty} e^{t(k-1)}(1-p)^{k-1} \\ &= pe^t \sum_{k=1}^{\infty} (e^t(1-p))^{k-1} \\ &= pe^t \frac{1}{1-e^t+pe^t} \end{aligned}$$

Note that the sum converges for $e^t(1-p) < 1$. ■

Example 9.10.6
The Bernoulli process $b(p)$ has $M_X(t) = 1-p+pe^t$.

Proof. Let $X \sim b(p)$. X takes two values 1 with probability p and 0 with probability $q = 1 - p$. Then

$$M_X(t) = E(\exp(tX)) = p \cdot \exp(t \cdot 1) + (1-p) \cdot \exp(t \cdot 0) = pe^t + 1 - p.$$

■

Corollary 9.10

For a Bernoulli process $b(p)$,

$$M_X(t) = E(\exp(tX)) \leq \exp(p(e^t - 1)).$$

Proof. Starting with $M_X(t) = E(\exp(tX)) = pe^t + 1 - p$, we rewrite $pe^t + 1 - p = 1 + p(e^t - 1)$ and apply $e^x \geq 1 + x$ to obtain $1 + p(e^t - 1) \leq \exp(p(e^t - 1))$, as needed. ■

9.11 Indicator random variable

Definition 9.54 **Indicator Random Variable**

For an event A we define the indicator random variable I_A as follows.

- $I_A(s) = 1$ if $s \in A$, and
- $I_A(s) = 0$ if $s \notin A$.

The indicator random variable sometimes uses 1 for I. It is also known as the characteristic function of A.

Theorem 9.25

Let $X = I_{A_1} + \ldots + I_{A_n}$. Then

$$E(X) = E(I_{A_1} + \ldots + I_{A_n}) = \sum_i P(A_i).$$

Example 9.11.1

The number of fixed points on a random permutation p on $\{1, \ldots, n\}$ is one. We define a random variable A with

$$A(p) = |\{i : p(i) = i\}|$$

Then we generate $A_i(p) = 1$ if $p(i) = i$ and 0 otherwise. Then $A(p) = \sum_i A_i(p)$.

$$E(A_i) = P[p(i) = i] = 1/n.$$

and

$$E(A) = \sum_i E(A_i) = n \cdot 1/n = 1.$$

9.12 Characteristic function

Definition 9.55

Let X be any random variable. The characteristic function c_X of X is defined as follows

$$c_X(t) = E(e^{itX}),\ t \in \mathbb{R}.$$

Note that $i = \sqrt{-1}$. Moreover,

$$c_X(t) = E(\cos(tX) + i\sin(tX)),\ t \in \mathbb{R}.$$

Theorem 9.26

Let X be any random variable with finite first k moments. Then the following apply.

$$\begin{aligned} c_X(0) &= 1 \\ c_X'(0) &= i \cdot E(X) \\ c_X'''(0) &= -i \cdot E(X^2) \\ &\dots \\ c_X^{(m)}(0) &= i^m \cdot E(X^m). \end{aligned}$$

9.13 Conditional expectation

Definition 9.56

For a discrete r.v. X and an event A with $P(A) > 0$, the condiional expectation of X given A is given by the following expression.

$$E(X|A) = \sum_x xP(X = x|A) = \sum_x x \cdot \frac{P(X = x, A)}{P(A)}.$$

Consider a die with $A = \{2,4,6\}$. Let X be the outcome of rolling a die. $P(X = 3|A) = 0$ but $P(X = 4|A) = 1/3$.

Definition 9.57

For two discrete r.v. X, Y with $P(Y = y) > 0$, the condiional expectation of X given $Y = y$ is given by the following expression.

$$E(X|Y = y) = \sum_x xP(X = x|Y = y) = \sum_x x \cdot \frac{P(X = x, Y = y)}{P(Y = y)} = \sum_x x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Definition 9.58

For two discrete r.v. X, Y the condiional expectation of X given Y is a random variable $E(X|Y)$ which is equal to $E(X|Y = y$ for $Y = y$.

Definition 9.59 **Conditional expectation: continuous case**

For two continuous random variables X and Y, with joint density $f_{X,Y}(x,y)$ the conditional expectation of X given Y is given below. Note that $f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y)$.

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} x f_{X,Y}(x,y)dx.$$

Definition 9.60

For three discrete r.v. X, Y, Z, let A be an event and a, b, z be real numbers. Moreover let $W = aX + bY$. Then the following apply.

1. $E(W|A) = aE(X|A) + E(Y|A)$.
2. $E(W|Z = z) = aE(X|Z = z) + bE(Y|Z = z)$.
3. $E(W|Z) = aE(X|Z) + bE(Y|Z)$.

Similar definitions apply for absolutely continuous random variable analogs of the variable in the previous definitions.

Definition 9.61

For two r.v. X, Y with joint density function $f_{X,Y}(x,y)$ the conditional expectation of X given $Y = y$ is given by the following expression.

$$E(X|Y = y) = \int x f_{X|Y}(x|y)dx = \int x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)}dx.$$

Example 9.13.1

Let X,Y be two r.v with $f_{X,Y}(x,y) = x^2+y^2$ for $0 \le x,y \le 1$ and 0 otherwise. For $0 < y < 1$ we have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y} dx = \int_0^1 (x^2+y^2)dx = x^3/3 + y^2 x\Big|_0^1 = y^2 + 1/3.$$

Then

$$E(X|Y=y) = \int x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx == \int_0^1 x \frac{x^2+y^2}{y^2+1/3} dx = \frac{3+6y^2}{12y^2+4}.$$

Definition 9.62

For two r.v. X,Y that are (absolutely) continuous, the condiional expectation of X given Y is a random variable $E(X|Y)$ which is equal to $E(X|Y=y$ for $Y=y$.

Definition 9.63

If X,Y are random variables then

$$E(E(X|Y)) = E(X).$$

Theorem 9.27

If X,Y are random variables and $g : \mathbb{R} \mapsto \mathbb{R}$ then

$$(a) E(E(X|Y)g(Y)) = E(Xg(Y)),$$

and

$$(b) E(g(Y)X|Y) = g(Y)E(X|Y),$$

and

$$(c) E(E(X|Y)|Y) = E(X|Y).$$

Proof. (a) The first result is proven as follows for the discrete case. The continuous case proof is similar.

$$\begin{aligned}
E(E(X|Y)g(Y)) &= \sum_y g(y)E(X|Y=y)P(Y=y) \\
&= \sum_y g(y)\left(\sum_x xP(X=x|Y=y)\right)P(Y=y) \\
&= \sum_y g(y)\left(\sum_x x\frac{P(X=x,Y=y)}{P(Y=y)}\right)P(Y=y) \\
&= \sum_y g(y)\sum_x xP(X=x,Y=y) \\
&= \sum_x\sum_y xg(y)P(X=x,Y=y) \\
&= \sum_x\sum_y xg(y)f_{X,Y}(x,y) \\
&= E(g(Y)X).
\end{aligned}$$

(b) The second result is proven as follows for the discrete case. The continuous case proof is similar.

$$\begin{aligned} E(g(Y)X|Y) &= E(g(Y)X|Y=z)\forall z \\ &= \sum_x \sum_y g(y)xP(X=x,Y=y|Y=z) \\ &= \sum_x g(y)xP(X=x|Y=z) \\ &= g(y)\sum_x xP(X=x|Y=z) \\ &= g(Y)E(X|Y=z)\forall z. \end{aligned}$$

Because $E(g(Y)X|Y=z) = g(Y)E(X|Y=z)$, $\forall z$ then $E(g(Y)X|Y) = g(Y)E(X|Y)$.
(c) The third result is proven as follows for the discrete case. The continuous case proof is similar.

$$\begin{aligned} E(E(X|Y)|Y) &= E(E(X|Y)|Y=z)\forall z \\ &= \sum_y E(X|Y=y)P(Y=y|Y=z) \\ &= E(X|Y=z)\forall z. \end{aligned}$$

Because $E(E(X|Y)|Y=z) = E(X|Y=z)$, $\forall z$ then $E(E(X|Y)|Y) = E(X|Y)$. ■

9.14 Probability inequalities

Definition 9.64 **Convex function**

Let $f(x)$ be a real valued function on some interval $[a,b]$ with $b \geq a$. Function f is called convex if for all $x_1 \neq x_2 \in [a,b]$, and for all λ such that $0 \leq \lambda \leq 1$ we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

In other words, a function f is convex if the line segment between any two distinct points on the graph of the function lies above or on the graph between the two points.

Equivalently, A differentiable function of one variable is convex on an interval if and only if its graph lies above all of its tangents i.e. $f(x) \geq f(y) + f'(y)(x-y)$. If function f on one variable is twice differentiable, then f is convex if $f''(x) \geq 0$.

9.14.1 Markov's inequality

Theorem 9.28

For a non-negative r.v. X with expectation μ that is thus non negative and every positive $a > 0$ the following applies.

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof.

$$Z = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$$

It is clear that $Z \leq X$. Therefore $E(Z) \leq E(X)$.

$$E(Z) = a \cdot P(Z = a) + 0 \cdot P(Z = 0) = a \cdot P(Z = a) = aP(X \geq a).$$

By way of $E(X) \geq E(Z) \geq aP(X \geq a)$, the result follows.
Another proof considers an Indicator function, and deals with the intrinsic case above.

$$I = \left\{ \begin{array}{ll} 0 & X < a \\ 1 & X \geq a \end{array} \right\}$$

Then

$$P(X \geq a) = E[I] \leq E(\frac{X}{a}) \leq \frac{1}{a}E(X).$$

■

A traditional proof for the discrete and continuous cases works as follows.
(a) **Discrete case**. If X is non negative then $X(s) \geq 0$ and we then obtain the following.

$$\begin{aligned} E(X) &= \sum_{X(s)} X(s)p(s) \\ &= \sum_{0 \leq X(s) < a} X(s)p(s) + \sum_{X(s) \geq a} X(s)p(s) \\ &\geq \sum_{X(s) \geq a} X(s)p(s) \\ &\geq a \cdot \sum_{X(s) \geq a} p(s) \\ &= a \cdot P(X(s) \geq a) = a \cdot P(X \geq a). \end{aligned}$$

(b) **Continuous case**.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_{-\infty}^{a} x f_X(x)dx + \int_{a}^{\infty} x f_X(x)dx \\ &\geq \int_{a}^{\infty} x f_X(x)dx \\ &\geq a \cdot \int_{a}^{\infty} f_X(x)dx \\ &\geq a \cdot \left(1 - \int_{-\infty}^{a} f_X(x)dx\right) \\ &\geq a \cdot (1 - P(X \leq a)) \\ &\geq a \cdot P(X \geq a) \end{aligned}$$

Technically, the latter (last line bound) is $P(X > a)$. We could then rephrase the inequality as $P(X > a) \leq E(X)/a$.

9.14.2 Cauchy-Schwartz inequality

Theorem 9.29

For any r.v. X,Y, assuming $Var(X),Var(Y)$ exist, are non-zero, and are finite, the following applies.

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}.$$

If $Var(Y) > 0$ equality above is for $X - E(X) = a(Y - E(Y))$, with $a = Cov(X,Y)/Var(Y)$.

Proof.
If $Var(Y) = 0$ then since $Var(Y) = E(Y - E(Y))^2) \geq 0$ we have $Y = E(Y)$ with probability 1. Then

$$Cov(X,Y) = E((X - E(X))(Y - E(Y)) = E((X - E(X))(E(Y) - E(Y)) = 0 = \sqrt{Var(X) \cdot 0}.$$

Assume now that $Var(Y) \neq 0$ (i.e. $Var(Y) > 0$). Set $S = X - E(X)$, $T = Y - E(Y)$.

$$\begin{aligned} E((S - aT)^2) &= E(S^2) + a^2 E(T^2) - 2aE(ST) \\ &= E((X - E(X))^2) + a^2 E((Y - E(Y))^2) - 2aE(ST) \\ &= Var(X) + a^2 Var(Y) - 2aCov(X,Y) \end{aligned}$$

The left-hand size $E((S - aT)^2) \geq 0$ for all a. Thus the right-hand side should also be so for all a. Therefore the discriminant of the corresponding quadratic equation must be ≤ 0. Therefore

$$4Cov(X,Y)^2 - 4Var(X)Var(Y) \leq 0 \Rightarrow Cov(X,Y)^2 \leq Var(X)Var(Y) \Rightarrow |Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$$

Moreover $|Cov(X,Y)| = \sqrt{Var(X)Var(Y)}$ when the discriminant is zero which implies that the quadratic equation has one real root. There is an a such that $E((S - aT)^2) = 0$ and therefore $S = aT$ with probability one. Then

$$Cov(X,Y) = E(ST) = E(aT^2) = aE(T^2) = aVar(Y),$$

and thus $a = Cov(X,Y)/Var(Y)$ for $Var(Y) > 0$. ■

Another formulation with a more compact proof follows.

Theorem 9.30 **Cauchy-Schwartz inequality**

Let A and B be two random variables. with $E(A^2) < \infty$ and $E(B^2) < \infty$. Then

$$|E(AB)| \leq \sqrt{E(A^2)}\sqrt{E(B^2)}.$$

Proof. Consider $E((A - tB)^2) \geq 0$. Then

$$E((A - tB)^2) = E(A^2) + t^2 E(B^2) - 2tE(A)E(B) \geq 0.$$

Choose $t = \frac{E(AB)}{E(B^2)}$. Then we have

$$\begin{aligned} E(A^2) + \left(\frac{E(AB)}{E(B^2)}\right)^2 E(B^2) - 2\frac{E(AB)}{E(B^2)}E(AB) &\geq 0 \\ E(A^2) + \frac{E^2(AB)}{E(B^2)} - 2\frac{E^2(AB)}{E(B^2)} &\geq 0 \\ E(A^2) - \frac{E^2(AB)}{E(B^2)} &\geq 0 \\ E(A^2)E(B^2) &\geq E^2(AB). \end{aligned}$$

The last inequality, after taking square roots concludes the proof. ■

9.14.3 Tsebyshev's or Tchebychef's inequality

From now use we use the more frequently spelled out Chebyshev.

Theorem 9.31 **Chebyshev's inequality**

Let X be a random variable X with finite expected value μ and non-zero and finite standard deviation σ (variance σ^2). Then for every positive $t > 0$,

$$P(|X-\mu| \geq \sigma t) \leq \frac{1}{t^2}.$$

Proof. It derives directly from Markov's inequality. Note that in Markov's case μ must be non-negative. This is not needed here since the use of $(X-\mu)^2$ implies an expected value that is non-negative.

$$P(|X-\mu| \geq \sigma t) = P(|X-\mu|^2 \geq \sigma^2 t^2) \leq E[(X-\mu)^2]/s^2t^2 = \sigma^2/s^2t^2 = 1/t^2.$$

■

An alternative formulation follows, if one replaces σt with t.

Theorem 9.32 **Tsebyshev's inequality**

For any r.v. X with $t > 0$, assuming $E(X)$ exists and is finite, the following applies.

$$P(|X-E(X)| \geq t) \leq \frac{Var(X)}{t^2}.$$

Proof. Consider $Y = (X-E(X))^2$. Use Markov's inequality on Y.

$$P(|X-E(X)| \geq t) = P(|X-E(X)|^2 \geq t^2) \leq \frac{E(X-E(X))^2}{t^2} \leq \frac{Var(X)}{t^2}.$$

■

Note that for any $m > 0$ there is a r.v. Z with the inequality being turned into an equality. For example consider r.v. Z with $P(Z=t) = 1/2 = P(Z=-t)$. Then $E(Z) = 0$ and $Var(Z) = t^2$, and $P(|Z| \geq t) = 1$.

The following form starts from $P(|X| \geq t)$ instead of $P(|X-\mu| \geq t)$ or $P(|X-\mu| \geq \sigma t)$.

Corollary 9.11 **Chebyshev variation 2**

An alternative form is the following bound.

$$P(|X| \geq t) \leq \frac{E(X^2)}{t^2}.$$

And of course be reminded $var(X) = \sigma^2$. One can obtaine sharper bounds for small variance by considering $P(|X-\mu|^k \geq \sigma t)$, where $k > 2$.

If we repeat an experiment n times we can consider the outcome of each experiment a random variable and thus talk about $X_1, X_2, \ldots, X_n$, where X_i is the random variable associated with the i-the outcome. If the experiments are independent of each other, then $P(X_i = x_k, X_j = x_l) = P(X_i = x_k)P(X_j = x_l)$.

Corollary 9.12

Let S be a random variable that is defined as $S = X_1 + \ldots + X_n$, where X_i are independent random variables. Then $var(S) = \sum_i var(X_i)$. Chebyshev's inequality is then given in the forn below

$$P\left(\frac{1}{n}|\sum_{i=1}^{n}(X_i - E[X_i])| \geq t\right) \leq \frac{\sigma^2}{nt^2},$$

where $\sigma^2 = (1/n)\sum_i var(X_i)$.

Proof. For S as defined, let $E[S] = \sum_i E[X_i]$.

$$\begin{aligned} P(|S - E[S]| \geq nt) &\leq \frac{var(S)}{n^2t^2} \\ P(\frac{1}{n}|S - E[S]| \geq t) &\leq \frac{var(S)}{n^2t^2} \\ P(\frac{1}{n}|\sum_i X_i - E[\sum_i X_i]| \geq t) &\leq \frac{n\sigma^2}{n^2t^2} \\ P(\frac{1}{n}|\sum_i (X_i - E[X_i])| \geq t) &\leq \frac{\sigma^2}{nt^2} \end{aligned}$$

■

Corollary 9.13

Let X_i, $i = 1, \ldots, n$, be n (independent) Bernoulli trials with $X_i \sim b(1/2)$. (X_i are 0 or 1.) Let $S = \sum_i X_i$ as before. Show that $Pr(S \geq 3n/4) \leq 4/n$.

Proof. We have that $E[X_i] = 1/2$ and so is $E[X_i^2] = 1/2$. Thus $var(X_i) = (1/2) - (1/2)^2 = 1/4$. For $S = \sum_i X_i$, we have $var(S) = nvar(X_i) = n/4$, and $E[S] = n/2$. Applying Chebyshev's inequality,

$$P(S \geq 3n/4) = P(|S - E[S]| \geq n/4) \leq var(S)/(n/4)^2 = 4/n.$$

■

Definition 9.65 **Sample Mean, Average**

Let X be a r.v. with mean μ and standard deviation σ. Let an experiment is repeated n time with repetitions independent of each other. Let X_i be the random variable associated with the i-th repetition. Then the sample mean or average of $X_1, \ldots, X_n$ is defined as follows.

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

The sample mean $\bar{X}$ is also a random variable.

Theorem 9.33 **Law of Large Numbers**

Let X_i, $i = 1, \ldots, n$, be n independent trials with finite expected value $E(X_i) = \mu$ and finite variance $Var(X_i) = \sigma^2$. Let

$$S_n = X_1 + X_2 + \ldots + X_n.$$

Then for any $\varepsilon > 0$ we have the following

$$P(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon) \to 0,$$

as $n \to \infty$ and equivalently,

$$P(\Big|\frac{S_n}{n} - \mu\Big| \leq \varepsilon) \to 1,$$

as $n \to \infty$.

A proof follow's by Chebyshev's inequality for $X = S_n/n$, and $E(X) = E(S_n)/n = n\mu/n = \mu$ and $Var(X) = Var(S_n/n) = (1/n^2)Var(S_n) = (1/n^2)nVar(X_i) = \sigma^2/n$.

In other words, For any $\varepsilon > 0$ the probablity that the sample mean $\bar{X}$ of n independent experiments has a value in the interval $[\mu - \varepsilon, \mu + \varepsilon]$ is to the limit for large n equal to 1.

$$\lim_{n\to\infty} P(|X - \mu| \leq \varepsilon) = 0.$$

9.14.4 Jensen's inequality

Theorem 9.34

For any r.v. X, let $f : \mathbb{R} \mapsto \mathbb{R}$ be a convex function with $E(f(X))$ finite. Then

$$f(E(X)) \leq E(f(X)).$$

Equality above is for $f(X) = aX + b$ for some $a, b \in \mathbb{R}$. If $f(x) = x^2$, this provided $(E(X))^2 \leq E(X^2)$, which has alread been derived.

9.14.5 Holder's inequality

Theorem 9.35

For any r.v. X,Y, and $a,b \in (1,\infty)$ satisfy $1/a+1/b=1$, the following applies.

$$E(|XY|) \leq (E(|X|^a)^{1/a}(E(|Y|^b)^{1/b}.$$

For $a=b=2$ the following form of the Cauchy-Schartz inequality is derived.

$$E(|XY|) \leq \sqrt{(E(X^2)(E(Y^2)}.$$

9.14.6 Liapunov's inequality

Theorem 9.36

For any r.v. X,Y, and $a < b \in (1,\infty)$, the following applies.

$$E(|X|^a)^{1/a} \leq (E(|X|^b)^{1/b}.$$

9.14.7 Minkowski's inequality

Theorem 9.37

For any r.v. X,Y, and $c \in [1,\infty)$, the following applies.

$$E(|X+Y|^a)^{1/a} \leq (E(|X|^a)^{1/a} + (E(|X|^a)^{1/a}.$$

For $a = 1$ the triangular inequality is derived.

Definition 9.66

For three discrete r.v. X,Y,Z, let A be an event and a,b,z be real numbers. Moreover let $W = aX + bY$. Then the following apply.

1. $E(W|A) = aE(X|A) + E(Y|A)$.
2. $E(W|Z = z) = aE(X|Z = z) + bE(Y|Z = z)$.
3. $E(W|Z) = aE(X|Z) + bE(Y|Z)$.

9.14.8 Hoeffding inequality

Proposition 9.2 **Hoeffding [23], Eq (4.16)**

For a random variable X such that $a \leq X \leq b$ with $E(X) = 0$ and for any $t > 0$, we have the following.

$$E[e^{tX}] \leq \exp\left(\frac{t^2(b-a)^2}{8}\right).$$

Proof. The function $f(x) = \exp(tx)$ is a convex function. Therefore

$$e^{tX} \leq \frac{x-a}{b-a}e^{tb} + \frac{b-x}{b-a}e^{ta}.$$

Moreover, $E(X) = 0$. By taking expectations on both sides we conclude the following.

$$\begin{aligned} E(e^{tX}) &\leq E(\frac{x-a}{b-a}e^{tb} + \frac{b-x}{b-a}e^{ta}) \\ &\leq \frac{E(x)-a}{b-a}e^{tb} + \frac{b-E(x)}{b-a}e^{ta} \\ &\leq \frac{0-a}{b-a}e^{tb} + \frac{b-0}{b-a}e^{ta} \\ &\leq \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta} \end{aligned}$$

Consider

$$E(e^{tX}) \leq \exp\left(\lg\left(\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}\right)\right) = \exp(\ln(g(t))).$$

The function $g(t)$ will be rewritten as $f(t) = \ln(g(t))$, and furthermore by some change of variables, using $p = b/(b-a)$ and therefore $1-p = -a/(b-a)$, and $x = (b-a)t$. Then we have, after renaming variables with substitution, the following. Note that $ta = (x/(b-a)) \cdot a = (x/(b-a)) \cdot (p-1)(b-a) = x(p-1)$.

$$\begin{aligned} f(t) &= \ln(g(t)) \\ f(t) &= \ln\left(\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}\right) \\ &= \ln\left(e^{ta}\left(\frac{-a}{b-a}e^{t(b-a)} + \frac{b}{b-a}\right)\right) \\ &= \ln\left(e^{ta}\left((1-p)e^{t(b-a)} + p\right)\right) \\ &= \ln\left(e^{ta}\left((1-p)e^{x} + p\right)\right) \\ &= ta + \ln\left((1-p)e^{x} + p\right) \\ f(x) &= x(p-1) + \ln\left((1-p)e^{x} + p\right) \end{aligned} \tag{9.1}$$

Function $f(x)$ as indicated in Eq 9.1 has the following properties: $f(0) = 0$ and $f'(0) = 0$. We point out that

$$f'(x) = p - 1 + \frac{(1-p)e^x}{p + (1-p)e^x},$$

and

$$f''(x) = \frac{(1-p)e^x(p + (1-p)e^x) - (1-p)e^x(1-p)e^x}{(p + (1-p)e^x)^2} = \frac{p(1-p)e^x}{(p + (1-p)e^x)^2}.$$

Using Taylor's formula we obtain

$$f(x) = f(0) + f'(0) + f''(r)x^2/2!,$$

for some r. We are going to find the maximum of $f''(x)$. We observe that $f''(x) = A \cdot B$, where $A + B = 1$, with $A = (p)/(p + (1-p)e^x)$ and $B = ((1-p)e^x)/(p + (1-p)e^x)$. Therefore the second derivative is maximized for $A = B = 1/2$ and $f''(x) \leq 1/4$. This implies.

$$f(x) = f(0) + f'(0) + f''(r)x^2/2! \leq 0 + 0 + (1/4)(1/2)x^2 \leq x^2/8.$$

Moving backwards, Equation (9.1) then yields, after recovering the orignal variable names,

$$\begin{aligned} E(e^{tX}) &\leq \exp(\ln(g(t))) \\ &\leq \exp x^2/8 \\ &\leq \exp\left((b-a)t^2/8\right). \end{aligned}$$

■

9.15 More on expectations

Theorem 9.38

Let X be a discrete r.v. with values $x_1, x_2, \ldots$ and $p_i = P(X = x_i)$. Then

$$E(X) = \sum_i x_i p_i = \int_0^\infty P(X > t)dt - \int_{-\infty}^0 P(X < t)dt,$$

assuming the two integrals $\int_0^\infty P(X > t)dt < \infty$, $\int_{-\infty}^0 P(X < t)dt < \infty$.

Proof. We separate the positive x from the negative x's. (In other words assume below that all x's are positive, and then all x's are negative: a combination of the two cases yields the final result.) For the positive case.

$$\begin{aligned}
\sum_i p_i x_i &= \sum_i p_i \int_0^{x_i} dt \\
&= \int_0^\infty \sum_{i, x_i > t} p_i dt \\
&= \int_0^\infty P(X > t).
\end{aligned}$$

For the negative case.

$$\begin{aligned}
\sum_i p_i x_i &= -\sum_i p_i \int_{x_i}^0 dt \\
&= -\int_{-\infty}^0 \sum_{i, x_i < t} p_i dt \\
&= -\int_{-\infty}^0 P(X < t).
\end{aligned}$$

The combination yields the result. ∎

Theorem 9.39

Let X be an absolutely continuous r.v. with p.d.f $f_X(x)$. Then

$$E(X) = \int_{-\infty}^\infty f_X(x)dx = \int_0^\infty P(X > t)dt - \int_{-\infty}^0 P(X < t)dt,$$

assuming the two integrals $\int_0^\infty P(X > t)dt < \infty$, $\int_{-\infty}^0 P(X < t)dt < \infty$.

Proof. Similarly to the discrete case we work as follows. For the positive case.

$$\begin{aligned}
\int_0^\infty P(X > t)dt &= \int_0^\infty (\int_t^\infty f_X(x)dx)dt \\
&= \int_0^\infty (\int_0^x f_X(x)dt)dx \\
&= \int_0^\infty (x f_X(x))dx.
\end{aligned}$$

For the negative case.

$$\begin{aligned}\int_{-\infty}^{0} P(X<t)dt &= \int_{-\infty}^{0}(\int_{-\infty}^{t} f_X(x)dx)dt \\ &= \int_{-\infty}^{0}(\int_{x}^{0} f_X(x)dt)dx \\ &= \int_{-\infty}^{0}(-xf_X(x))dx \\ &= \int_{-\infty}^{0}(-xf_X(x))dx.\end{aligned}$$

The combination yields the result. ■

Theorem 9.40

For $1 \leq i \leq n$, let X_i be a random variable with c.d.f. $F_{X_i}(x)$. Let Z be a r.v. that is a combination of the X_i so that for $\sum_i a_i = 1$, and $a_i > 0$ (or ≥ 0), $F_Z(z) = \sum a_i F_{X_i}(x)$. Then

$$E(Z) = \sum_i a_i E(X_i).$$

Proof. Consider $P(Z > z)$.

$$\begin{aligned}P(Z>z) &= 1 - P(Z<z) \\ &= 1 - F_Z(z) \\ &= 1 - \sum a_i F_{X_i}(x) \\ &= \sum_i a_i - \sum a_i F_{X_i}(x) \\ &= \sum_i a_i P(X_i > z).\end{aligned}$$

One can show a similar result for $P(Z < z)$. Then

$$\begin{aligned}E(Z) &= \int_0^{\infty} P(Z>z)dz - \int_{-\infty}^{0} P(Z<z)dz \\ &= \int_0^{\infty} \sum_i a_i P(X_i>z)dz - \int_{-\infty}^{0} \sum_i a_i P(X_i<z)dz \\ &= \sum_i a_i E(X_i).\end{aligned}$$

■

9.16 Spaces

Exercise 653.
A (S,T) is a topological space where $S=\{a,b,c\}$ and $T=\{\emptyset,S,\{a,b\},\{b,c\},\{b\}\}$. But it is not a σ-algebra. Explain.

Solution.
The complement of $\{a,b\}$ is not in T for example. ■

Exercise 654.
A finite intersection of open intervals $\cap_{m\in\{1,3,5\}}(a-\frac{1}{m},b+\frac{1}{m})$ preserves the open interval notion. How about a countable intersection of intervals?

Solution.
No. $\cap_{m\in N}(a-\frac{1}{m},b+\frac{1}{m})$ is a closed interval $[a,b]$. Consider also the case $a=b=0$ or $a=0,b=1$.

■

Exercise 655.
(a) For a probability space (S,T,P) show that $P(\emptyset)=0$.
(b) For a probability space (S,T,P) show that for all events $A\in T$,

$$P(S-A)=1-P(A).$$

Solution.
Note that $S\cup\emptyset=S$ and $S\cap\emptyset=\emptyset$. Thus S an $\emptyset$ are obviously disjoint. Since P is a probability measure we have

$$S=S\cup\emptyset\Rightarrow P(S)=P(S)+P(\emptyset)$$

Give that $P(S)=1$ by the definition of a probability measure, solving for $P(\emptyset)$ we obtain the obvious $P(\emptyset)=0$.
Note that $S-A\cap A=\emptyset$. Moreover $(S-A)\cup A=S$. Therefore

$$S=(S-A)\cup A\Rightarrow P(S)=P(S-A)+P(A)\Rightarrow 1=P(S-A)+P(A),$$

and we conclude the proof. ■

9.17 Events and their properties

Exercise 656.

For an event A we define the indicator function I_A as follows.

- $I_A(s) = 1$ if $s \in A$, and
- $I_A(s) = 0$ if $s \notin A$.

The indicator function sometimes uses 1 for I. It is also known as the characteristic function or indicator function of A.

Later on we might refer to it as the indicator random variable. The indicator function is also known as the characteristic function.

(a) For any event A, we have

$$E(I_A) = P(A).$$

(b) For any two events A, B we have for all $s \in S$, the following.

$$I_{A\cap B}(s) = I_A(s)I_B(s), \quad I_{A\cup B}(s) = I_A(s) + I_B(s) - I_{A\cap B}(s), \quad I_{A'}(s) = 1 - I_A(s).$$

Solution.

$$E(I_A) = \sum_{s\in S} I_A(s)p(s) = \sum_{s\in A} I_A(s)p(s) + \sum_{s\notin A} I_A(s)p(s) = \sum_{s\in A} I_A(s)p(s) + \sum_{s\notin A} 0\cdot p(s) = P(A)$$

For every $s \in S$ we have if $s \in A \cap B$ then $s \in A$ and $s \in B$. Therefore $I_{A\cap B}(s) = I_A(s) = I_B(s) = 1$. Then $I_{A\cap B}(s) = I_A(s)I_B(s) = 1$. If $s \notin A \cap B$ then either $s \notin A$ or $s \notin B$. First $s \notin A \cap B$ implies $I_{A\cap B}(s) = 0$ and one of $I_A(s)$, $I_B(s)$ is also zero. Therefore $I_{A\cap B}(s) = 0 = I_A(s) \cdot I_B(s)$. The result is proven. The other two cases can be proven similarly. ■

Exercise 657.
Let $A, B \in T$ and $A \subseteq B$. Then $P(A) \leq P(B)$.

Solution.
Consider A and $B-A$. The two are disjoint since $A \subseteq B$. The following then applies.

$$P(A \cup (B-A)) = P(B) \Leftrightarrow P(A) + P(B-A) = P(B) \Leftrightarrow P(B-A) = P(B) - P(A).$$

Since $P(B-A) \geq 0$ then $P(B) - P(A) \geq 0$ i.e. $P(A) \leq P(B)$. ■

Exercise 658.

Inclusion-exclusion. Let (S,T,P) be a probability space and $A,B \in T$. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Solution.
Let for each $s \in S$ we define $p_s = P(\{s\})$. Then utilizing the previous problem we obtain the following.

$$\begin{aligned} P(A \cup B) &= \sum_{s \in S} p_s I_{A \cup B}(s) \\ &= \sum_{s \in S} p_s I_{A \cup B}(s) \\ &= \sum_{s \in S} p_s (I_A(s) + I_B(s) - I_{A \cap B}(s)) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Another way to prove this result is by applying the three axioms of Kolmogorov. Let $C = A \cap B$, $D = A - B$ and $F = B - A$. Note that the three sets are mutually disjoint. Moreover $A \cup B = C \cup D \cup F$.

$$\begin{aligned} P(A) + P(B) - P(A \cap B) &= (P(C) + P(D)) + (P(C) + P(F)) - P(C) \\ &= P(C) + P(D) + P(F) \\ &= P(A \cup B) \end{aligned}$$

■

Exercise 659.
Show that for any collection of events $A_1, \ldots, A_n$,

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n) = \sum_{i=1}^{n} P(A_i).$$

Solution.
Consider, events B_i such that

$$B_i = A_i - (A_1 \cup \ldots \cup A_{i-1})$$

Then $\cup_i B_i = \cup A_i$ and $P(B_i) \leq P(A_i)$ and the events B_i are disjoint. By additivity of the probability measure we have

$$P(A_1 \cup \ldots \cup A_n) = P(B_1 \cup \ldots \cup B_n) = \sum_i P(B_i) \leq \sum_i P(A_i)$$

■

Alternatively, for two events, we use the previous problem. Then induction can be used.

Exercise 660.

Let S be a sample space and $A_i \subseteq S$ events. For each non empty subset $I \subseteq \{1,\ldots,n\}$ we define

$$A_I = \cap_{i\in I} A_i.$$

By default $A_0 = S$.
We have

$$|A_1 \cup \ldots \cup A_n| = \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|A_I|. = \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|\cap_{i\in I} A_i|.$$

Solution.
By induction on n. For $n=1$, trivially $|A_1| = |A_1|$. From n to $n+1$ we use the $n=2$ case.

$$\begin{aligned}
|\cup_{i=1}^{n+1} A_i| &= |(\cup_{i=1}^{n} A_i) \cup A_{n+1}| \\
&= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |(\cup_{i=1}^{n} A_i) \cap A_{n+1}| \\
&= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |\cup_{i=1}^{n} (A_i \cap A_{n+1})| \\
&= \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|\cap_{i\in I} A_i| + |A_{n+1}| - \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|(\cap_{i\in I} A_i \cap A_{n+1})| \\
&= \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|\cap_{i\in I} A_i| + |A_{n+1}| - \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|(\cap_{i\in I\cup\{n+1\}} A_i)| \\
&= \sum_{I\subseteq\{1,\ldots,n+1\}, n+1\notin I} (-1)^{|I|+1}|\cap_{i\in I} A_i| + \sum_{I\subseteq\{1,\ldots,n+1\}, n+1\in I} (-1)^{|I|+1}|(\cap_{i\in I} A_i)| \\
&= \sum_{I\subseteq\{1,\ldots,n+1\}} (-1)^{|I|+1}|\cap_{i\in I} A_i|
\end{aligned}$$

■

9.18 Toys: dies, coins and card

9.18.1 Throw or roll a die or dice

Exercise 661.

The singular form of dice is die. A die has six faces associated with six possible numbers, one on each face of a die. Each number is indicated by an appropriate number of black filled dots.
An experiment is performed by rolling a (pair of) dice and the number indicated at the top of the dice (opposite to the base that sits on a surface) is recorded. Give the sample space S and calculate its cardinality.

Solution.
The sample space S has 36 outcomes

$$S = \{(a,b) : 1 \leq a,b \leq 6\}$$

■

Exercise 662.

Throw or roll? We again toss a pair of dice. Sample space S contains 36 ordered pairs (a,b) as elements where a shows the number (top face) of one die and b shows the number of the other die during a toss.

$$S = \{(a,b) : 1 \leq a,b \leq 6\}$$

Random variable X is defined as follows

$$X : S \to \mathbb{R} \quad \text{where } X(s) = X((a,b)) = a+b, \quad s \in S.$$

Thus for an element $s = (a,b)$ of S, random variable X assigns a value to this element that is equal to the sum of the numbers of the faces of the two dice.
What is the range of X?
Random variable Y is defined as follows

$$Y : S \to \mathbb{R} \quad \text{where } Y((a,b)) = \min(a,b).$$

Thus for an element $s = (a,b)$ of S, random variable Y assigns a value to this element that is equal to the minimum of the two values of the two faces of the two dice.
What is the range of Y?

Solution.
The range of X is $R(S,X) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$.
The range of Y is $R(S,Y) = \{1,2,3,4,5,6\}$. ■

Exercise 663.

We have two dice. The sample space of rolling two dice is of cardinality 36, as there are so many pairs (d_1, d_2) where d_1 is one of the outcomes of one die, and likewise for d_2. Use the definitions to calculate $P(X = 4)$, where X is a random variable defined to be the sum $d_1 + d_2$.

Solution.
It is clear that $P(X) = 1/12$. We properly define for $A \subseteq \mathbb{R}$, the set $X^{-1}(A) = \{s \in S : X(s) \in A\}$, that is the set of outcomes of S that result in getting mapped by X to a value in A. This implies,

$$P(X \in A) = P(X^{-1}(A)).$$

The set $X^{-1}(A)$ must be measurable and thus it should belong to T i.e. $X^{-1}(A) \in T$. Say $A = \{4\}$. Then

$$X^{-1}(A) = \{(1,3), (2,2), (3,1)\}$$

Therefore

$$P(X \in A) = P(\{(1,3), (2,2), (3,1)\}) = 3/36 = 1/12.$$

■

Exercise 664.
(a) We roll a die six times. How many outcomes?
(b) We roll a die six times. What is the probability that all outcomes of the six rolls are distinct from each other?

Solution.
(a) $6 \times 6 \times 6 \times 6 \times 6 \times 6$.
(b) $6!/6^6$. ■

9.18.2 Toss a coin

Exercise 665.

All coins hereafter are fair coins unless otherwise stated.
An experiment is performed by throwing (tossing) a coin. The experiment is repeated twice. The combination of the results of the two individual experiments is the experiment in question.
What is the sample space S?

Solution.
The sample space S is then $S = \{HH, HT, TH, TT\}$ and indicates the outcomes of the first and second toss of the coin: H indicates heads, T indicates tail as an outcome. AB indicates that A is the outcome of the first experiment (toss of the coin), B is the outcome of the second experiment (second toss of the coin), where A, B is H or T. Define event X, where $X = \{HH, TT\}$ i.e. even number of H or T. Likewise define event Y, where $Y = \{HH, HT, TH\}$ i.e. at least one H. ■

Exercise 666.
An experiment is performed by tossing a coin. The experiment is repeated until an H is encountered. Is the sample space finite or infinite?

Solution.
The sample space is infinite. Why? Because $S = \{T, TH, TTH, TTTH, TTTTH, \ldots\}$. ■

Exercise 667.

Let p be a real number $0 \leq p \leq 1$. We toss a biased coin; p is the probability that coin comes H and thus $q = 1 - p$ that it comes T. The coin is tossed repeteadly until the the first H comes. The outcomes of such a sequence of coin tosses can be described by a variable length vector

$$\underbrace{(\overbrace{T,T,\ldots,T}^{k},H)}_{k+1},$$

of length $k+1$, where k is the length of T tosses. Such a sequence can be uniquely identified by the number k of tosses. We can thus define the sample space S to be $S = \mathbb{N}$. For each $k \in S$ we then define

$$P(\{k\}) = (1-p)^k p$$

Show that $P(.)$ is a probability measure.

Solution.

$$\sum_{k \in N} P(\{k\}) = \sum_{k \in N} (1-p)^k p = p \cdot \sum_{k \in N} (1-p)^k = p \frac{1}{1-(1-p)} = 1,$$

since $0 \leq p \leq 1$. ■

Exercise 668.

A coin is tossed twice. The number of heads is recorded. What is its sample space S?
What is the probability model assigned to S's elements?
For event $A = \{1,2\}$ what is $P(A)$?
For event $B = \{2\}$ what is $P(B)$?

Solution.
The sample space is $S = \{0,1,2\}$. The following probability model is assigned $p(0) = 1/4, p(1) = 1/2, p(2) = 1/4$.
Event A with $A = \{1,2\}$ has $p(A) = 3/4$, and event B with $B = \{2\}$ has $p(B) = 1/4$. ■

Exercise 669.
Flip a fair coin three times. The probability of H or T each time is $1/2$ respectively.
(a) What is the probability of having exactly one tail?
(b) What is the probability of having at least one head?

Solution.
(a) $3/8$: Out of the 8 outcomes only three maps two one tail: THH, HTH,HHT.
(b) $7/8$: The probability is 1 minus the probability of have three tails. The only such even TTT has probability $1/8$ thus $1-1/8$ is the answer. ■

Exercise 670.

When we toss a coin 4 times (as we did before), and we call this an experiment each toss is a subexperiment that is known as a trial. Thus this experiment has 4 identical trials. The two outcomes H and T can be characterized as success and failure and assigned probabilities p and q respectively with $p+q=1$. For a fair coin $p=q=1/2$. In a Bernoulli trial the two outcomes labelled success and failure (H and T respectively in this example) are also indicated with 1 for success and 0 with failure. In a Rademacher trial success is indicated with $+1/2$ and failure with $-1/2$.
Multiple Bernoulli trials can be independent of each other or not. They are usually the former.
A coin is tossed four times. What is the probability of having exactly two successes (i.e. two H or equivalently two ones)? If we add the labels of the outcomes this is equivalent of getting 2 (successes, heads) as well.

Solution.
There are 16 possible outcomes for this experiment.

$$HHHH, HHHT, HHTH, HHTT, HTHH, HTHT, HTTH, HTTT,$$

$$THHH, THHT, THTH, THTT, TTHH, TTHT, TTTH, TTTT.$$

Each such outcome occurs with probability $1/16$. 6 of those outcomes correspond to the event of exactly 2 Heads occurring. Thus the probability is $6 \cdot 1/16$. Equivalently

$$P(H=2) = \binom{4}{2} (1/2)^2 (1/2)^2$$

The first term indicates the possibility of getting two H, when we flip of coin 4 times, the second is the probability of having 2 Heads and the third one of having $4-2=2$ Tails T in such tosses. The tosses are indpependent. In general for k successes and $n-k$ failures in a toin coss that is repeated (independently) n times is as follows capturing also the possibility that the coin is not fair (p is the probability of success i.e. H, and $q=1-p$ of failure): the probability of k successes and $n-k$ failures is $p^k q^{n-k}$. But there are $C(n,k) = \binom{n}{k}$ ways of having k successes in n trials. Thus $P(n,k) = C(n,k) p^k q^{n-k}$. Note that by the binomial theorem,

$$\sum_{k=0}^{n} P(n,k) = \sum_{k=0}^{n} C(n,k) p^k q^{n-k} = (p+q)^n = 1.$$

■

Exercise 671.
Toss a coin $n = 2m$ times. What is the probability of exactly $n/2 = m$ heads?

Solution.
Let $p_k = \binom{n}{k} p^k (1-p)^{n-k}$ Assuming the coin is fair $p = q = 1-p = 1/2$ and thus $n = 2m$ and $n/2 = m$, we have

$$p = \binom{n}{k} p^k (1-p)^{n-k} \binom{2m}{m} p^m (1-p)^{2m-m} = \binom{2m}{m} (1/2)^m (1/2)^{2m-m} = \binom{2m}{m} (1/2)^{2m}. = \binom{n}{n/2} (1/2)^n.$$

Furthermore we use Stirling's formula for $n > 10$ i.e. $n! \approx \sqrt{2\pi n}(n/e)^n$

$$p = \frac{n!}{(n/2)!(n/2)!} \frac{1}{2^n} = \frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi n/2}(n/2e)^{n/2} \sqrt{2\pi n/2}(n/2e)^{n/2}} \frac{1}{2^n} = \frac{\sqrt{2}}{\sqrt{\pi n}}$$

■

Exercise 672.

You have a fair coin where in a coin toss the probability of Heads is $p_H = 1/2$ and so is the probability of Tails $p_T = 1/2$. (From now on we use H or T to indicate one or the other outcome.)
(a) What is the expected number of tosses for a T? (For example, in the experiment HHHT we have four tosses to a T.)
(b) What is the expected number of tosses to get TH? (For example, in the experiment TTTH we have four tosses to a TH.)

Solution.
(a) It is a geometric distribution with expected number of tosses equal to $1/p_T$, where $p_T = 1/2$, i.e. two tosses.
(b) First in an expected number of two tosses i.e. $1/p_T$ we get a T. From that point on in an expected number of $1/p_H$ tosses we get an H. By the sum of expectations $E[A+B] = E[A] + E[B]$ we get that the answer is the sum $1/p_T + 1/p_H = 4$.

■

Exercise 673.
We toss a fair coin n times. It comes H in each one of the n tosses. We toss it once more. What is the probability that it comes H?

Solution.
$1/2$. Implicit is the assumption that the tosses are independent of each other. Drawing any other conclusion or using any either biased and thus incorrect information is also not correct.
That is, H is not more likely to occur because it has already occurred n times : therefore we should not use the biased assumption that the toss is not fair and favors H, thus making sense to bet on H again. Furthermore we should not assume T is now more likely to occur : for a fair coin to land H n times in a row is an indicator that next time is the time for a T. Past history gets wiped out priot to the next toss of fair coin toss of a fair coin! ■

Exercise 674.

We roll a die three times. What is the probability (a) all outcomes are distinct (three distinct values), (b) we get two distinct values, (c) we get one distinct value, (d) what is the expected number of distinct outcomes?

Solution.
Let N be the number of distinct outcomes.

$$P(N=1)=6/6^3=1/36, \quad P(N=3)=6\cdot 5\cdot 4/6^3=5/9, \quad P(N=2)=1-P(N=1)-P(N=2)=1-1/36-20/36=15/36.$$

The

$$E(N)=1\cdot(1/36)+2\cdot(15/36)+3\cdot(5/9)=91/36.$$

■

9.18.3 Deck of cards

Exercise 675.

Deck of cards. A deck of cards consists of 52 cards. There are 4 suits known as clubs(C), diamonds(D), hearts(H), and spades(S). The hearts and diamonds are red and spades and clubs are black. Each suit contains 13 cards numbered 2 through 10, three face cards, jack (J), queen (Q), and king (K), and ace (A) for one.
From a deck of cards we select one card c. We define two events A and B as follows.

$$A = \{c \textbf{ is diamond } \}, \qquad B = \{c \textbf{ is a face card } \}.$$

Compute $P(A)$, $P(B)$, $P(A \cap B)$.

Solution.
$P(A) = 13/52 = 1/4$. $P(B) = (3*4)/52 = 3/13$. $P(A \cap B) = 3/52$. ■

Exercise 676.
(a) Partition a deck of cards among 4 people each one getting 13 cards.
(b) What is the probability each person gets a King?

Solution.

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{(13!)^4}.$$

Or

$$(a1a2\ldots a12a13)(b1b2\ldots b12b13)(c1c2\ldots c12c13)(d1d2\ldots d12d13)$$

Set 52 place holders $a\ldots b\ldots c\ldots d\ldots$ and parentheses on a line. We have 52! to arrange the cards of the deck in each group. Within a group order does not matter, thus we divide by 13! for each group Total $52!/(13!)^4$.
(b) Partition the King first in $4\cdot 3\cdot 2\cdot 1$ ways. And then the remaining 48 cards into the 4 players in $48!/(12!)^4$ ways. Combining (a) and (b) so far we have the following

$$\frac{4\cdot 3\cdot 2\cdot \left(48!/(12!)^4\right)}{52!/(13!)^4}.$$

■

9.19 Conditional probabilities

Exercise 677.
For two events

$$P(X|Y) = P(X \cap Y)/P(Y).$$

Furthermore

$$P(X \cap Y) = P(X|Y)P(Y) = P(Y|X)P(X).$$

If X and Y are independent events, then

$$P(X \cap Y) = P(X)P(Y).$$

This implies $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$.
(a) Let $P(A) = 1/2$ and let $P(A \cap B) = 1/8$. Find the probability $P(B/A)$.
(b) A bin contains 8 balls, three of them are black and the remaining red. A person draws a ball and then draws a second ball from the bin. What is the probability that are BOTH black if there is no replacement of the first ball, after it is drawn.
(c) Same problem as in (b). Now the ball gets reinserted into the bin after the drawing.

Solution.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/8}{1/2} = \frac{1}{4}.$$

Let B be the event the first ball is black. $P(B) = 3/8$. Let C be the event the second ball is black as well. Then $P(C/B) = 2/7$. Therefore,

$$P(B \cap C) = 3/8 \cdot 2/7 = 6/56.$$

Let B be the event the first ball is black. $P(B) = 3/8$. Let C be the event the second ball is black as well. Then $P(C/B) = 3/8$. Therefore,

$$P(B \cap C) = 3/8 \cdot 3/8 = 9/64.$$

■

Exercise 678.
(a) Let A, B be two events with $0 < P(B) < 1$. Then show the following.

$$P(A) = P(A|B)P(B) + P(A|B')P(B').$$

(b) In a class of 50 kindergarten students 30 are boys and 20 are girls. 10 of the boys and 10 of the girls are allergic to peanuts. A child is drawn of the 50 students. What is the probability $P(A)$ that is allergic to peanuts?

Solution.
(a) Note that

$$A = A \cap S = A \cap (B \cup B') = (A \cap B) \cup (A \cap B').$$

The two events $(A \cap B)$ and $(A \cap B')$ are disjoint (mutually exclusive). Therefore

$$P(A) = P(A \cap B) + P((A \cap B').$$

Use the definition of conditional probability to conclude the proof.
(b) Let A be the even that a child is allergic to peanuts. We would like to computer $P(A)$. Let B be the event that a child is a boy. Then $P(B) = 30/50$. Moreover $P(G) = P(B') = 20/50$. Furthermore $P(A|B) = 10/30$ and $P(A|B') = 10/20$. We use the formula from part (a) to derive the following.

$$P(A) = P(A|B)P(B) + P(A|B')P(B'). = (10/30)(30/50) + (10/20)(20/50) == 10/50 + 10/50 = 0.4$$

■

Exercise 679.
Let (S,T,P) be a probability space. Let B be an event of it with $0 < P(B) < 1$, $B \in T$. Then define (S,T,P') as follows.

$$P'(A) = P(A|B), \quad \forall A \in T.$$

Show that P' is a probability measure.

Solution.

$$P'(S) = P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Moreover,

$$\begin{aligned} P'(A) &= P(A|B) = \frac{P(A \cap B)}{P(B)} \\ &= \frac{1}{P(B)} \sum_{s \in A \cap B} P(\{s\}) \\ &= \sum_{s \in A \cap B} P(\{s\}|B) \\ &= \sum_{s \in A} P(\{s\}|B) \\ &= \sum_{s \in A} P'(\{s\}), \end{aligned}$$

as needed. Note that

$$\sum_{s \in A} P(\{s\}|B) = \sum_{s \in A \cap B} P(\{s\}|B) + \sum_{s \in A \cap B'} P(\{s\}|B) = \sum_{s \in A \cap B} P(\{s\}|B) + \sum_{s \in A \cap B'} 0 = \sum_{s \in A \cap B} P(\{s\}|B).$$

. ■

Exercise 680.
Show that $P(A_1|A_2) = \frac{P(A_2|A_1)P(A_1)}{P(A_2)}$.

Solution.
Note that $P(A_2|A_1)P(A_1) = P(A_1 \cap A_2)$. Moreover, $P(A_1|A_2)P(A_2) = P(A_1 \cap A_2)$. ■

Exercise 681.
Let S be a sample space, $A \in T$ and $S_1, \ldots, S_m$ be a partition of S. Show the following.

$$P(A) = \sum_{i=1}^{m} P(A|S_i)P(S_i).$$

Solution.

$$A = A \cap S = A \cap (\cup_i S_i) = \cup_{i=1}^{m} (A \cap S_i).$$

Then the following applies.

$$P(A) = P(\cup_{i=1}^{m} (A \cap S_i)) = \sum_i P(A \cap S_i),$$

Since S_i are pairwise independent and so are $A \cap S_i$ subsets of them. We have that $P(A \cap S_i) = P(A|S_i)P(S_i)$, and the result follows. ■

9.20 Independence of events

Exercise 682.
(Repetition of definition.) Two events A, B are independent if

$$P(A \cap B) = P(A) \cdot P(B).$$

Show that if A and B are independernt, so are A and B', the latter being the complement of B.
Conclude then that A' and B' are also independent.

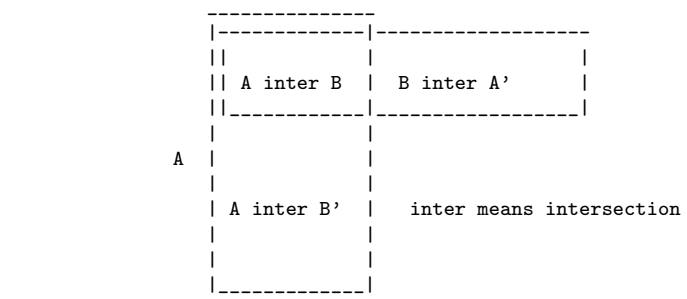
Solution.
Note that $A = (A \cap B) \cup (A \cap B')$. Then

$$P(A) = P(A \cap B) + P(A \cap B') = P(A)P(B) + P(A \cap B').$$

Solving for $P(A \cap B')$ we obtain

$$P(A \cap B') = P(A)(1 - P(B)) = P(A)P(B'),$$

and the result follows.



Repeat the same for A and B' now. ■

Exercise 683.

For three events (or more), A,B,C are pairwise independent if and only if

$$P(A\cap B)=P(A)P(B),P(B\cap C)=P(B)P(C),P(A\cap C)=P(A)P(C).$$

For three events (or more), A,B,C are mutually independent if and only if

$$P(A\cap B)=P(A)P(B),P(B\cap C)=P(B)P(C),P(A\cap C)=P(A)P(C).P(A\cap B\cap C)=P(A)P(B)P(C)$$

In general, for n events $A_1,\ldots,A_n$, they are mutually independent if and only if for every k, $k\leq n$ and for every k indices $i_1<i_2<\ldots i_k\leq n$

$$P(\cap_{j=1}^{k}A_{i_j})=\prod_{j=1}^{k}P(A_{i_j}).$$

Thus the condition is a formula that it is true for all 2^n-1-n subsets of those n events.
Consider throwing two independent coin tosses. The coin is fair. Let event A be first toss is H. Let event B be second toss is H. Let event C be two tosses are identical.
(a) Are A,B,C pairwise independent ?
(b) Are A,B,C independent ?

Solution.
(a) Clearly $P(A)=1/2$ and so is $P(B)=1/2$. For C, we have $C=\{HH,TT\}$ and thus $P(C)=1/2$. We thus obtain the following.

$$P(A)=P(B)=P(C)=1/2,P(A\cap C)=P(A\cap B)=1/4,P(B\cap C)=P(B\cap A)=1/4,$$

and therefore

$$P(A\cap C)=P(A)P(C),P(A\cap B)=P(A)P(B),P(B\cap C)=P(B)P(C),$$

implying A,B,C are pairwise independent. That is A,C are independent, and so are B,C and of course A and B.

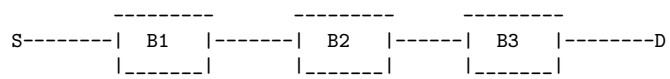
(b) We calculate $P(A\cap B\cap C)=1/4$. But $P(A)P(B)P(C)=(1/2)(1/2)(1/2)=1/8$. Therefore

$$P(A\cap B\cap C)\neq P(A)P(B)P(C).$$

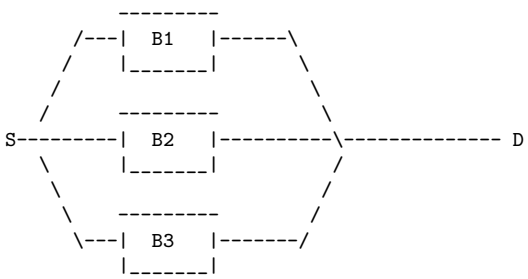
■

Exercise 684.

A passage from S to D consists of three bridges. They are raised bridges and have probability of failure to be down (and passable) of q_1, q_2, q_3 and thus they work ok with $p_1 = 1 - q1, p_2 = 1 - q_2, p_3 = 1 - q_3$.
(a) Serial setup. The bridges are set one after the other. What is the probability of failure to reach D from S? Assume bridges are working independently of one another.



(b) Parallel setup. The bridges are set up in parallel. What is the probability of failure to reach D from S? Assume bridges are working independently of one another.



Solution.
(a) Let OK1, OK2, OK3 be event that bridge B1, B2, B3 is passable. Let OK be the event that we can move from S to D. Thus P(FAIL) = 1 - P(OK).

$$P(FAIL) = 1 - P(OK) = 1 - P(OK1 \cap OK2 \cap OK3) = 1 - p_1 p_2 p_3$$

(b) Let OK1, OK2, OK3 be event that bridge B1, B2, B3 is passable. Let OK be the event that we can move from S to D. Thus P(FAIL) = 1 - P(OK).

$$\begin{aligned} P(FAIL) &= 1 - P(OK) = 1 - P(OK1 \cup OK2 \cup OK3) \\ &= 1 - (1 - P(FAIL1 \cap FAIL2 \cap FAIL3)) \\ &= P(FAIL1)P(FAIL2)PFAIL3 \\ &= q_1 q_2 q_3 = (1 - p_1)(1 - p_2)(1 - p_3). \end{aligned}$$

■

Exercise 685.
A (pair of) dice is rolled. The probability of any elementary event (outcome) of S is $1/36$.
Let A be the event a dice is 5. What is $P(A)$?
Let B be the event the sum of the dice is 6. What is $P(B)$?
What is $P(A|B)$?. This reads probability the event A occurs given that B has already occurred.

Solution.
$P(A) = 1/6$. We define B as follows.

$$B = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

It follows that $P(B) = 5/36$. The probability that a dice is 5 if the sum is 6 is $2/5$. Then $A|B$

$$A|B = \{(1,5),(5,1)\}$$

Obviously $P(A|B) = 2/5$. Moreover $P(A \cap B) = 2/36$. And it follows that

$$P(A|B) = P(A \cap B)/P(B) = (2/36)/(5/36) = 2/5.$$

. ■

Exercise 686.
We toss a coin 8 times. Assume it is a fair coin i.e. $P(H) = p = 1/2$ and thus $P(T) = (1-p) = 1/2$.
(a) What is the probability we have 3 heads? Call this event A.
(b) What is the probability we have 3 tails?
(c) What is the probability the first two tosses give an H each, and the remaining ones one H only for a total of 3 H? Call this event B.
(d) Call event C that the first two tosses give an H. What is the probability $P(C|A)$?

Solution.
(a) $\binom{n}{3}p^3(1-p)^{n-3}$ with $n = 8$ and $p = 1/2$.
(b) $\binom{n}{3}p^{n-3}(1-p)^3$ with $n = 8$ and $p = 1/2$. For $p = 1/2$ it is identical to the result of (a).
(c) $p \cdot p \cdot \binom{n-2}{1}p^1(1-p)^{n-2-1}$ with $n = 8$ and $p = 1/2$.
(d)

$$P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{P(B)}{P(A)} = \frac{p \cdot p \cdot \binom{n-2}{1}p^1(1-p)^{n-2-1}}{\binom{n}{3}p^3(1-p)^{n-3}} = \frac{n-2}{\binom{n}{3}}.$$

■

Exercise 687.

We roll a die twice. Let A be the event that the first roll is a three. Let B be the event that the second roll is odd. Let C be the event that the sum of the two rolls is a six.
(a) How much is $P(A)$, $P(B)$, and $P(A \cap B)$?
(b) Are A, B independent?
(c) How much is $P(C)$.
(d) Are A, C independent?

Solution.
(a) For S we have

$$S = \{(i, j) : 1 \leq i, j \leq 6\}.$$

For event A we have the following.

$$A = \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)\}.$$

Therefore $P(A) = |A|/|S| = 6/36 = 1/6$. For event B we have the following.

$$B = \{(i,1), (i,3), (i,5) : 1 \leq i \leq 6\}.$$

Therefore $P(B) = |B|/|S| = 18/36 = 1/2$. For $P(A \cap B)$ we have the following.

$$A \cap B = \{(3,1), (3,3), (3,5)\}.$$

Therefore $P(A \cap B) = 3/36 = 1/12$.
(b) We note the following

$$P(A \cap B) = 1/12 = (1/6) \cdot (1/2) = P(A) \cdot P(B).$$

Therefore A, B are independent.
(c) For event C we have the following.

$$C = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}.$$

For event $A \cap C$ we have the following.

$$A \cap C = \{(3,3)\}.$$

Therefore $P(A \cap C) = 1/36$.
(d) We note the following

$$P(A \cap C) = 1/36 \neq (1/6) \cdot (1/36) = P(A) \cdot P(C).$$

Therefore A, C are not independent. ■

Exercise 688.

Out of a deck of 52 cards (four suits of thirteen cards) we pick two cards without replacement.
(a) What is the cardinality of S?
(b) Let event A be the event that the second card is a diamond. What is $P(A)$?
(c) Let event B be the event that the second card is a diamond given that the first card is a spade. What is $P(B)$?
(d) Let event C be the event that the first card is a spade. For event $A|C$, calculate $P(A|C)$.

Solution.
(a) $|S| = 52 \cdot 51$.
(b) There are two possibilities for A. Both card picked are diamonds. Or the first card picked is NOT a diamond, but the second card picked for A is indeed a diamond. Let us call the events A_1, A_2 respectively Thus A_1 is the event (x,y) where both x,y are diamonds. And A_2 is the even (z,y), where z is anything but a diamond and y is diamond. Obviously

$$|A_1| = 13 \cdot 12, \quad |A_2| = (52-13) \cdot 13.$$

$A_1 \cap A_2 = \emptyset$, and therefore

$$P(A) = \frac{|A_1| + |A_2|}{52 \cdot 51} = \frac{13 \cdot 12 + 13 \cdot 39}{52 \cdot 51} = 13/52 = 1/4.$$

(c) We now calculate $P(B)$.

$$B = \{(s,d) : 1 \leq s \leq 13, 1 \leq d \leq 13\}.$$

Then

$$P(B) = \frac{13 \cdot 13}{52 \cdot 51} = 13/204.$$

(d) We note that $|C| = 13$ and thus $P(C) = 13/52 = 1/4$. Moreover $P(A \cap C)$ is the probability the first card is a spade and the second a diamond. The cardinality of $A \cap C$ is then $|A \cap C| = 13 \cdot 13$ and thus $P(A \cap C) = 13 \cdot 13/(52 \cdot 51)$. Therefore,

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{(13/52)(13/51)}{1/4} = \frac{13}{51}.$$

Thus the apriori $P(A) = 1/4 = 13/52$ has been slighly increased into the $P(A|C) = 13/51$ now that we have the knowledge that the first card is a spade. ■

Exercise 689.

Bayes theorem. For probability space (S,T,P), let A,B be two events, with $P(A),P(B)\neq 0$. Then show the following.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

Caution: Some books use / instead of |. This is because A/B then means $A-B$ rather than $A|B$.

(a) Then deduce the following.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}.$$

In a population of 100,000 individuals, 15 of them have myocarditis. A test has been develope which is accurate 98% of the time. An individual takes the test and it says positive. What is the probability that the individual indeed has myocarditis?

Solution.

(a) Use $P(A\cap B)=P(B\cap A)$, and $P(A|B)=P(A\cap B)/P(B)$ and its analogue. From a prior result,

$$P(B)=P(B|A)P(A)+P(B|A')P(A'),$$

and the result follows.

(b) Let A be the event that an individual is infected with myocarditis, and let B be the event that the test comes positive. We need to compute $P(A|B)$. We thus use the following formula.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}.$$

We calculate $P(B|A)=0.98$ and thus $P(B|A')=0.02$. Moreover $P(A)=15/100000$ and thus $P(A')=99985/100000$. The result follows. ■

Exercise 690.
Two statistic professors, Alan and Bob, go out for pizza lunch. They decide who pays for the lunch by tossing a coin.
(a) They toss a fair coin thrice. If number of H is greater than number of T, Alan pays. Otherwise Bob pays. Who is more likely to pay?
(b) They toss the coin for the first time and it is H. Who is more likely to pay ?

Solution.
(a) None. By symmetry (swap H and T) they have the same probability of paying $1/2$.
(b) They toss the coin for the first time and it is H. The event A (first toss H) is then

$$A = \{HHH, HHT, HTH, HTT\}$$

Let Alan be the even alan pay and Bob be the event Bob pays. Then $P(Alan) = 3/4$, and $P(Bob) = 1/4$. This is because,

$$Alan = \{HHH, HHT, HTH\}, \quad Bob = \{HTT\}.$$

■

Exercise 691.
Show that two events A and B with $P(B) \neq 0$ are independent if and only if $P(A|B) = P(A)$.

Solution.
(a) If $A, B)$ are independent we know that $P(A \cap B) = P(A)P(B)$. Then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

(b) Moreover if $P(A|B) = P(A)$, then we have the following.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A).$$

The latter implies $P(A \cap B) = P(A)P(B)$, i.e. A and B are independent. Give that $P(B)$ appears in the denominator above, it explains the requirement that $P(B) \neq 0$. ■

Exercise 692.

This 4th of July, it is expected to have sunshine with probability 50%, to have cloudy weather with probability 25%, and to have rainy weather with probability 25%. An ice cream van expects to have a sell out in the former case with probability 75%, in the second case with probability 25%, and in the third case with probability 25% as well.

What is the probability of a sell out?

Solution.
Let S, C, R be the events sunshine, cloudy and rain accordingly.

$$P(S) = 0.5 \quad P(C) = 0.25 \quad P(R) = 0.25.$$

Let $P(T)$ be the probability of sell-out. We know the following

$$P(T|S) = 0.75, \quad P(T|C) = 0.25, \quad P(T|R) = 0.25.$$

We need to calculate $P(T)$.

$$P(T) = P(T|S)P(S) + P(T|C)P(C) + P(T|R)P(R) = 0.75 * 0.5 + 0.25 * 0.25 + 0.25 * 0.25 = 0.375 + 0.0625 + 0.0625 = 0.5.$$

■

Exercise 693.
In an urn we have two dimes, one quarter, and one nickel. We draw two coins without replacement.
(a) What is the probability that the first one is a dime.
(b) What is the probability that the second one is a dime.

Solution.
(a) The first one is a dime with probability $2/(2+1+1) = 1/2$.
(b) Let D be that the first draw was a dime, Q it was a quarter, and N a nickel. Then $P(D) = 2P(Q) = 2P(N) = 1/2$. Let S be the event the second draw is a dime.

$$P(S) = P(S|D)P(D) + P(S|Q)P(Q) + P(N|Q)P(N) = (1/3)(1/2) + (2/3)(1/4) + (2/3)(1/4) = 1/2.$$

■

Exercise 694.
For n events $A_1,\ldots,A_n$ assuming $P(A_1\cap\ldots\cap A_{n-1})\neq 0$ show than

$$P(A_1\cap\ldots\cap A_n)=P(A_n|A_1\ldots A_{n-1})\ldots P(A_2|A_1)P(A_1).$$

Use the result later for the Birthday problem.

Solution.
Use induction and work the case two for A_n and $A_1,\ldots,A_{n-1}$. ■

Exercise 695.

We toss a fair coin thrice (three times). Let A_1 be the even that the second and third tosses are the same. Let A_2 be the even that the first and third tosses are the same. Let A_3 be the even that the first two tosses are the same. (In other words A_i indicates the toss index i not used in the relevant definition.)
It is obvious that $P(A_1) = P(A_2) = P(A_3) = 1/2$.
Are A_1,A_2,A_3 mutually independent?

Solution.
NO. Consider $P(A_1 \cap A_2 \cap A_3)$. This describes the probability of the event where all coin tosses generate the same outcome (all H, or all T). Then, $P(A_1 \cap A_2 \cap A_3) = 2/8 = 1/4$. But $P(A_1)P(A_2)P(A_3) = 1/8 \neq P(A_1 \cap A_2 \cap A_3)$. Thus the three event are not mutually independent. ■

Exercise 696.
At NJIT passing CS 100 occurs with probability 0.75. (You repeat the course if you earned a grade of less than B.) Moreover if you fail the course for the fourth time you are ineligible to continue with the CS program. What is the probability of dicontinuing with the program? What is the probability of success.

Solution.
Let F indicates a failure to satisfy the requirement and S a success. The sample space S is

$$S = \{P, FP, FFP, FFFP, FFFF\}$$

The probability of discontinuing is $(1/4)^4 = 1/256 \approx 0.0039$. The probability of success is $1 - 1/256 > 0.996$. One can also compute the latter as follows

$$0.75 + 0.25 * 0.75 + (0.25)^2 0.75 + (0.25)^3 0.75 = 0.75(1 + 0.25 + 0.25^2 + 0.25^3) = 0.99609375.$$

■

9.21 Discrete random variables

Exercise 697.

Consider tossing a coin 10 times. The sample space S is vectors of length 10 that are binary: H for Heads or T for tails. Several times we simplify notation by using 1 for H, and 0 for T.

$$S = \{(a_1, a_2, \ldots, a_{10}) : a_i = 0, 1\}.$$

Let the coin be biased with p the probability for H (i.e. 1) and $1-p$ for the probability for T (0).
We ask the question. What is the probability of 3 H in the 10 coin tosses. In other words we are interested in $P(\sum_{i=1}^{10} a_i = 3)$.

Solution.
(a) We define random variable X

$$X = a_1 + a_2 + \ldots + a_{10} = \sum_{i=1}^{10} a_i,$$

from S to $\mathbb{N}$ i.e. $X : S \mapsto \mathbb{N}$. When we ask for $P(\sum_{i=1}^{10} a_i = 3)$ this is equivalent to $P(X = 3)$. Formally the latter is

$$P(X^{-1}(\{3\})).$$

We calculate

$$f_X(3) = P(X = 3) = \binom{10}{3} p^3 (1-p)^7.$$

■

Exercise 698.

Bernoulli distribution. Let (S,T,P) be a probability space. A 0-1 random variable Z follows the Bernoulli distribution with parameter p and we write $Z \sim b(p)$ if its mass function $f_Z : \{0,1\} \mapsto [0,1]$ satisfies $f_Z(1) = p$ and $f_Z(0) = q = 1-p$.

For a Bernoulli random variable Z the experiment is known sometimes as a trial and the event $Z = 1$ is identified as a success and the event $Z = 0$ as a failure.

Binomial distribution. Let (S,T,P) be a probability space. A $\{0,1,\ldots,n\}$ valued random variable X follows the Binomial distribution with parameters n,p and we write $X \sim B(n,p)$ if its mass function $f_X : \{0,1,\ldots,n\} \mapsto [0,1]$ satisfies

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for all k in $\{0,1,\ldots,n\}$. Moreover $\sum_{k=0}^{n} f_X(k) = 1$.

A binomial random variabl X counts the number of successes in n independent Bernoulli trials with parameter p.

(a) Reason part (a) of the previous problem in this context.

(b) For a Bernoulli random variable Z conclude $E(Z) = p$ and $Var(Z) = pq = p(1-p)$.

(c) For a Binomial random variable X conclude $E(X) = np$ and $Var(X) = npq = np(1-p)$.

Solution.

(a) The i-th toss of a coin follows a Bernoulli distribution. Thus $X_i = Z \sim b(p)$. The experiment with the 10 coin tosses follows a Binomial distribution. Let $X = X_1 + X_2 + \ldots + X_{10}$. Then $X \sim B(10,p)$. Formally, $X_i : S \mapsto \{0,1\}$. $X(s) = k$ if and only if k of the X_i's are 1 and the remaining 0. For each $V \subseteq [n] = [10] = \{0,1,\ldots,n\}$, $|V| = k = 3$, we define

$$A_V = \{s \in S : X_i(s) = 0 \ \forall i \notin V, X_i(s) = 1 \ \forall i \in V\}$$

Then

$$P(A_V) = p^3 (1-p)^{10-3},$$

since X_i are mutually independent. Sets $A_V : |V| = k = 3, V \subseteq \{0,1,\ldots,n\}$ partition $\{X = 3\}$ and therefore

$$f_X(3) = \sum_{V \subseteq [n], |V|=3} P(A_V) = \sum_{V \subseteq [n], |V|=3} p^3 (1-p)^{10-3} = \binom{10}{3} p^3 (1-p)^{10-3}.$$

The number of subsets of $[n]$ of cardinality 3 is $\binom{n}{3}$ and this justifies $\binom{10}{3}$ above. The result easily generalizes for n and k instead of 10 and 3.

(b)

$$E(Z) = 1 \cdot p + 0 \cdot q = p.$$

Moreover,

$$Var(Z) = 0^2 (1-p) + 1^2 p - p^2 = p(1-p) = pq.$$

(c) If X is the sum of n independent Bernoulli trials Z_i then we have the following.

$$E(X) = E(\sum_{i=1}^{n} Z_i) = \sum_i E(Z_i) = np.$$

Furthermore,

$$Var(X) = Var(\sum_{i=1}^{n} Z_i) = nVar(Z_i) = npq.$$

■

Exercise 699.

Geometric distribution.

Let X be a random variable that counts the number of trials to obtain the first success. Let the probability of success be $0 < p < 1$.
(a) Show that $E(X) = 1/p$.
(b) Show that $Var(X) = q/p^2$.

Solution.
(a) For $X = k$ we have $k-1$ failures before the first success. Therefore $P(X = k) = (1-p)^{k-1}p$. Then

$$E(X) = \sum_k k \cdot P(X = k) = \sum_k k \cdot (1-p)^{k-1}p = p\sum_k k \cdot (1-p)^{k-1}.$$

We note that

$$\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{1-(1-p)} \Rightarrow \left(\sum_{k=0}^{\infty}(1-p)^k\right)' = \left(\frac{1}{1-(1-p)}\right)' \Rightarrow -\sum_{k=0}^{\infty}k(1-p)^{k-1} = -\frac{1}{p^2}.$$

Combining the two we obtain the following.

$$E(X) = \sum_k k \cdot P(X = k) = p\sum_k k \cdot (1-p)^{k-1} = p/p^2 = 1/p.$$

(b)

$$Var(X) = E(X^2) - E(X)^2 = E(X^2) - 1/p^2$$

We work as before.

$$\begin{aligned}
\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{1-(1-p)} &\Rightarrow \\
\sum_{k=0}^{\infty}k(1-p)^{k-1} = \frac{1}{p^2} &\Rightarrow \\
\sum_{k=0}^{\infty}k(1-p)^{k} = \frac{1-p}{p^2} &\Rightarrow \\
\left(\sum_{k=0}^{\infty}k(1-p)^{k}\right)' = \left(\frac{(1-p)}{p^2}\right)' &\Rightarrow \\
\sum_{k=0}^{\infty}k^2(1-p)^{k-1}\cdot p = p\cdot\left(-\frac{(1-p)}{p^2}\right)' &\Rightarrow \\
\sum_{k=0}^{\infty}k^2(1-p)^{k-1}\cdot p = \frac{-(p-2)}{p^2}.&
\end{aligned}$$

For the latter $E(X^2)$ term we have the following.

$$E(X^2) = \sum_k k^2 \cdot (1-p)^{k-1} \cdot p = -(p-2)/p^2.$$

Therefore

$$Var(X) = E(X^2) - E(X)^2 = E(X^2) - 1/p^2 = -p/p^2 + 2/p^2 - 1/p^2 = (1-p)/p^2 = q/p^2.$$

■

Exercise 700.

Let X,Y be discrete r.v. and $a,b \in \mathbb{R}$. Show the following.

$$E(aX+bY) = aE(X)+bE(Y).$$

Conclude that for continuous X,Y the result also applies.

Solution.
Let X, Y be two discrete random variables taking also discrete values $x_1,\ldots,x_i,\ldots$ and $y_1,\ldots,y_j,\ldots$. The following derivations can be obtained. Let $Z = aX+bY$.

$$\begin{aligned}
E(Z) &= \sum_k z_k f_Z(z_k) \Leftrightarrow \\
E(Z) &= \sum_k z_k P(Z=z_k) \Leftrightarrow \\
E(aX+bY) &= \sum_k z_k P(aX+bY=z_k) \Leftrightarrow \\
E(aX+bY) &= \sum_i \sum_j \sum_{k,ax_i+by_j=z_k} z_k f_{X,Y}(x_i,y_j) \Leftrightarrow \\
E(aX+bY) &= \sum_i \sum_j (ax_i+by_j) P(X=x_i,Y=y_j) \Leftrightarrow \\
E(aX+bY) &= \sum_i \sum_j (ax_i+by_j) P(X=x_i,Y=y_j) \\
&= \sum_i \sum_j ax_i P(X=x_i,Y=y_j) + \sum_i \sum_j by_j P(X=x_i,Y=y_j) \\
&= a\sum_i x_i \sum_j P(X=x_i,Y=y_j) + b\sum_j y_j \sum_i P(Y=y_j,X=x_i) \\
&= a\sum_i x_i P(X=x_i) + b\sum_j y_j P(Y=y_j) \\
&= aE(X)+bE(Y).
\end{aligned}$$

Of course, utilizing Theorem 9.9 we can skip directly to step

$$E(aX+bY) = \sum_i \sum_j (ax_i+by_j) P(X=x_i,Y=y_j),$$

for $g(X,Y) = aX+bY$. For X and Y r.v. as before,

$$E(X+Y) = E(X)+E(Y),$$

obtainable by setting $a=b=1$. Similarly, the corresponding continuous case bound can be proven. Random variable X,Y must be jointly absolutely continuous, and $a,b \in \mathbb{R}$.

$$\begin{aligned}
E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax+by) f_{X+Y}(x,y)\,dx\,dy \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X+Y}(x,y)\,dx\,dy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X+Y}(x,y)\,dx\,dy \\
&= a\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X+Y}(x,y)\,dy\right)dx + b\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X+Y}(x,y)\,dx\right)dy \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} y f_Y(y)\,dy \\
&= aE(X)+bE(Y).
\end{aligned}$$

■

Exercise 701.
For any random variables $X_1, \ldots, X_n$ of a probability space, we have that

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i).$$

Solution.
By induction on n. (Number of random variables must be countable.) We alread proved it for $n = 2$, the base case with the previous example $X_1 = X$ and $X_2 = Y$. Inductive step follows trivially. ■

Exercise 702.

DISK DRIVE LATENCY. On the unit interval $[0,1]$ two points x,y are chosen uniformly at random. What is the expected distance of $|X-Y|$ of those two points?
On a hard disk drive we have N tracks (labeled 0 to $N-1$ or 1 to N). A hard-disk drive's head (or heads) moves, depending on track requests linearly from track i to track j covering a distance $|i-j|$. If the track requests are generated uniformly at random what is the average track movement of the head/heads?

Solution.
Let X,Y be the random variable representing the two points chosen. Then the joint p.d.f of X,Y is $f(x,y)=1$ for $0\leq x,y\leq 1$ and 0 otherwise.

$$\begin{aligned} E(|X-Y|) &= \int_0^1\int_0^1 |x-y|dxdy \\ &= \int\int_{x\geq y}(x-y)dxdy+\int\int_{x<y}(y-x)dxdy \\ &= \int_0^1\int_y^1 (x-y)dxdy+\int_0^1\int_0^y (y-x)dxdy \\ &= 1/6+1/6=1/3 \end{aligned}$$

■

Exercise 703.
For any random variables X,Y that are independent,

$$E(XY) = E(X) \cdot E(Y).$$

Solution.
Random variables X,Y are independent. Moreover they are jointly absolutely continuous. Therefore $f_{X,Y}(x,y) = f_X(x)f_Y(y) = P(X=x)P(Y=y)$. If X,Y are continuous given that they are independent, we have the following, utilizing the continuous version of Theorem 9.9.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} y f_Y(y)\int_{-\infty}^{\infty} x f_X(x)dx \\ &= E(X)E(Y). \end{aligned}$$

If X,Y are discrete and X,Y independent, $X \neq 0$, we have the following using Theorem 9.9 with $g(X,Y) = XY$.

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyP(X=x,Y=y) \\ &= \sum_x \sum_y xyP(X=x)P(Y=y) \\ &= \sum_x xP(X=x)\sum_y yP(Y=y) \\ &= E(X)E(Y). \end{aligned}$$

■

Exercise 704.

For two random variables X,Y with $X \leq Y$ show that $E(X) \leq E(Y)$. The implicit assumption is that X,Y share sample space. Therefore $X \leq Y$ more formally means the following

$$\forall s \in S : X(s) \leq Y(s).$$

Solution.
Consider random variable $Z = Y - X$. Then $Z(s) = Y(s) - X(s) \geq 0$ for all $s \in S$. We have the following for discrete X,Y,Z.

$$E(Z) = \sum_i z_i f_Z(z_i) = \sum_i z_i P(Z = z_i) \geq 0.$$

Since $E(Z) \geq 0$ and $E(Z) = E(Y - X) = E(Y) - E(X)$ we have as a sonsequence that $E(Y) - E(X) = E(Z) \geq 0$ and therefore $E(Y) \geq E(X)$.
For the continuous case, moreover X,Y are jointly absolutely continuous. Let $f_{X,Y}(x,y)$ be the joint probability density function of X and Y. Since $X \leq Y$ this implies that $f_{X,Y}(x,y) = 0$ for $x > y$. Consider random variable $Z = Y - X$. Then we have the following

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - x) f_{X,Y}(x,y) dx dy.$$

Since $f_{X,Y}(x,y) = 0$ for $x > y$ then $y - x \geq 0$ and of course $f_(X,Y)(x,y) \geq 0$. This implies $E(Z) \geq 0$ and furthermore $E(Z) = E(Y - X) = E(Y) - E(X) \geq 0$. We then obtain $E(Y) \geq E(X)$. ■

Exercise 705.
Let $n \in \mathbb{N}$ and $p \in [0,1]$ and let $X \sim B(n,p)$. Then show $E(X) = np$.

Solution.
Let $X = X_1 + \ldots + X_n$ be the sum of n Bernoulli trials. Then

$$E(X) = E(X_1 + \ldots + X_n) = \sum_{i=1}^{n} X_i = n \cdot E(X_i).$$

The X_i are identically distributed $X_i \sim b(p)$. Therefore

$$E(X_i) = p \cdot 1 + (1-p) \cdot 0 = p.$$

Therefore $E(X) = n \cdot p = np$. ■

Exercise 706.

Let X, Y be two random variables. Let $X \sim U(a,b)$. The drawing of real numbers uniformly at random in the closed interval $[a,b]$ can be represented by X. We would like to use probabilistic sorting to sort those real numbers but the algorithm works only for reals in the interval $[0,1]$. Map X into Y as needed.

Solution.

$$a \leq X \leq b \Rightarrow 0 \leq X-a \leq b-a \Rightarrow 0 \leq X-a \leq b-a \Rightarrow 0 \leq \frac{X-a}{b-a} \leq 1 \Rightarrow 0 \leq Y \leq 1.$$

■

9.22 Continuous random variables

Exercise 707.

Let X,Y be two random variables. Let $X \sim U(0,9)$. Let $Y = \sqrt{X}$
(a) Find the probability density function of X, and cumulative distribution function of X.
(a) Find the probability density function of X, and cumulative distribution function of Y.

Solution.
(a) $f_X(x) = 1/9$ and $F_X(x) = \frac{x-0}{9-0} = x/9$.
(b) If $X \sim U(0,9)$ then $Y = \sqrt{X} \sim U(0,3)$. Moreover $Y^2 = X$. Therefore

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(Y^2 \leq y^2) \\ &= P(X \leq y^2) \\ &= F_X(y^2) \\ &= y^2/9, \end{aligned}$$

for $0 \leq y \leq 3$. Note that $F_Y(y) = 0$ for $y < 0$, and $F_Y(y) = 1$ for $y > 3$, In order to computer $f_Y(y)$ we take the first derivative of $F_Y(y)$ with respect to y. Then $f_Y(y) = 2y/9$ for $0 \leq y \leq 3$, 0 otherwise.
In general if $Y = g(X)$, then $Y \leq y$ is equivalent to $g(X) \leq y$ or $X \leq g^{-1}(y)$.

$$\begin{aligned} Y \leq y, Y = g(X) &\Rightarrow g(X) \leq y \Rightarrow X \leq g^{-1}(y) \Leftrightarrow \\ F_Y(y) &= F_X(g^{-1}(y)) \\ \frac{dF_Y(y)}{dy} &= \frac{dF_X(g^{-1}(y))}{dy} \\ f_Y(y) &= f_X(g^{-1}(y)) \cdot (g^{-1}(y))' \end{aligned}$$

In the example $Y = g(X) = \sqrt{(}X)$ and therefore $X = g^{-1}(Y) = Y^2$. Then,

$$f_Y(y) = f_X(g^{-1}(y)) \cdot (g^{-1}(y))' = \frac{1}{9} \cdot 2y = 2y/9,$$

for $0 \leq y \leq 3$ since $f_X(x) = 1/9$ for $0 \leq x \leq 9$. ■

Exercise 708.
Let $X \sim U(0,9)$. Then $0 \leq X \leq 9$. Let $Y = X^2$. Then $0 \leq Y \leq 81$.
(a) Find $F_Y(y)$ and $f_Y(y)$.
(b) Find $E(X^2)$.

Solution.
(a)

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = \frac{\sqrt{y}}{9}.$$

This means

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{\sqrt{y}}{9} & 0 \leq y \leq 81 \\ 1 & y > 81. \end{cases}$$

Then

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{1}{18\sqrt{y}}.$$

This means

$$f_Y(y) = \begin{cases} 0 & y \leq 0 \\ \frac{\sqrt{1}}{18\sqrt{y}} & 0 \leq y \leq 81 \\ 0 & y > 81. \end{cases}$$

(b)

$$E(X^2) = \int_0^9 x^2 f_X(x)dx = \int_0^9 x^2 \frac{1}{9} dx = \frac{x^3}{27}\Big|_0^9 = 27.$$

■

Exercise 709.

Let now $X \sim N(0,1)$ and consider $Y = X^2$. Find the p.d.f of Y.

Solution.
We have that for X, $f_X(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. Then $F_X(x) = P(X \leq x)$. Moreover,

$$F_X'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

For $Y = X^2$, $Y \leq y$ is equivalent to $X^2 \leq y$ i.e. $-\sqrt{y} \leq X \leq \sqrt{y}$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= F_X(\sqrt{y}) - (1 - F_X(\sqrt{y})) \\ &= 2F_X(\sqrt{y}) - 1. \end{aligned}$$

Then

$$\begin{aligned} f_Y(y) &= \frac{d}{dY} F_Y(y) \\ &= 2F_X'(\sqrt{y}) \cdot (\sqrt{y})' \\ &= 2F_X'(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi y}} e^{-y/2}. \end{aligned}$$

For $y > 0$, $f_Y(y)$ is given by the expression above. For $y < 0$, $f_Y(y) = 0$. ■

Exercise 710.
(a) Let X,Y be two independent random variables $X \sim U(0,1)$ and $Y \sim U(0,1)$. Let $Z = \max\{X,Y\}$. Find the c.d.f. of Z.
(b) Then do the same for $W = \min\{X,Y\}$.
(c) What is the median of Z?

Solution.
For X,Y we have

$$F_X(x) = F_Y(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

If $Z = \max\{X,Y\} \le x$, the $X \le x$ and $Y \le x$. Moreover $P(X \le x) = P(Y \le x) = x$. (Note that X and Y are also independent.) The

$$P(Z \le x) = P(X \le x) \cdot P(Y \le x) = x^2.$$

Therefore,

$$F_Z(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

(b) For $W = \min\{X,Y\}$ we similarly have that $W \ge x$ if and only if $X \ge x$ and $Y \ge x$. Therefore,

$$P(W \ge x) = P(X \ge x) \cdot P(Y \ge x) = (1 - P(X \le x)(1 - P(Y \le x) = (1 - F_X(x))(1 - F_Y(x))$$

$$\begin{aligned} P(W \le x) &= 1 - P(W \ge x) \\ &= 1 - (1 - F_X(x))(1 - F_Y(x)) \\ &= 1 - (1 - F_X(x))^2 \\ &= 2x - x^2 \end{aligned}$$

(c) The median of Z is the value z_m such that $F_Z(z_m) = 1/2$, in other words $z_m^2 = 1/2$ i.e. $z_m = \sqrt{2}/2$. ■

Exercise 711.
Consider random variable X with probability density function

$$f_X(x) = \begin{cases} 1 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Consider random variabl X with probability density function

$$f_Y(y) = \begin{cases} 0.5 & x = 2 \\ 0.5 & x = 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $E(X)$ and $E(Y)$.
(b) Find $Var(X)$ and $Var(Y)$.

Solution.
(a) $E(X) = 1 \cdot 3 = 3$. $E(Y) = 0.5 \cdot 2 + 0.5 \cdot 4 = 3$.
(b)

$$Var(X) = E((X - E(X))^2) = E((3-3)^2) = 0.$$

Moreover,

$$Var(X) = E((X - E(X))^2) = (2-3)^2 \cdot 0.5 + (4-3)^2 \cdot 0.5 = 0.5 + 0.5 = 1$$

■

Exercise 712.
(a) For $X \sim U(0,1)$ find $f_X(x)$, $E(x)$ and $F_X(x)$.
(b) Repeat (a) for $X \sim U(a,b)$.

Solution.
(a)

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$E(x) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 1 dx = \frac{x^2}{2}\Big|_0^1 = 1/2.$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t) dt = \int_0^x f_X(t) dt = \int_0^x 1 dt = x.$$

Note that $F_X(x) = 0$ for $x < 0$, and Note that $F_X(x) = 1$ for $x > 1$.
(b)

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Then

$$E(x) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{x^2}{2} \cdot \frac{1}{b-a}\Big|_a^b = \frac{a+b}{2}.$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t) dt = \int_a^x f_X(t) dt = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$

Note that $F_X(x) = 0$ for $x < a$, and Note that $F_X(x) = 1$ for $x > b$. ■

Exercise 713.
Let $X \sim E(\lambda)$ so that

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(a) Find $E(X)$.

Solution.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_0^{\infty} x\lambda e^{-\lambda x} dx \\ &= -\int_0^{\infty} x de^{-\lambda x} \\ &= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 + \int_0^{\infty} e^{-\lambda x} dx \\ &= \frac{1}{-\lambda} e^{-\lambda x}\Big|_0^{\infty} \\ &= e^{-\lambda \cdot 0}/\lambda = \frac{1}{\lambda}. \end{aligned}$$

■

Exercise 714.
Let $X \sim N(0,1)$. Find $E(X)$.

Solution.
We have that for X, $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Then $F_X(x) = P(X \leq x)$. Moreover,

$$F_X'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Then

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x e^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x e^{-x^2/2} dx \end{aligned}$$

Function $xe^{-x^2/2}$ and the limits of the two summand integrals are symmetric. Then the two terms of $E(X)$ cancel out and thus $E(X) = 0$ (implied by $X \sim N(0,1)$).
A more elegant approach is to use moment generating functions $f(t) = E(e^{tX})$ for $X \sim N(0,1)$.

$$\begin{aligned} E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{tx - x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{(x-t)^2} dx \\ &= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{t^2/2} \int_{-\infty}^{\infty} e^{(x-t)^2/2} dx \\ &= e^{t^2/2}. \end{aligned}$$

The second term is the sum of $f_Y(y)$ where $Y \sim N(t,1)$ and thus the integral is 1, and the last term is obtained. Furthermore

$$e^{t^2/2} = E(e^{tX}) = 1 + tE(X)/1! + t^2 E(X^2)/2! + \ldots$$

Take the first derivative of the left-hand side and the right-hand side. Then equate $t = 0$. One then obtains $E(X) = 0$. ■

Exercise 715.

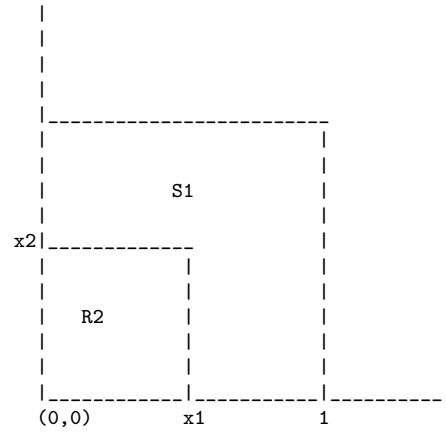
Let (x_1,x_2) be a point chosen uniformly at random in the unit square with coordinates $(0,0),(1,0),(1,1),(0,1)$ counterclock-wise. Form r.v. $X=x_1^2$, $Y=x_2^2$ and $Z=x_1+x_2$. Show that X,Y are independent but Y,Z are not by determining the $F_{X,Y}(x,y)$ and $F_{Y,Z}(x,z)$.

Solution.
We note the following.

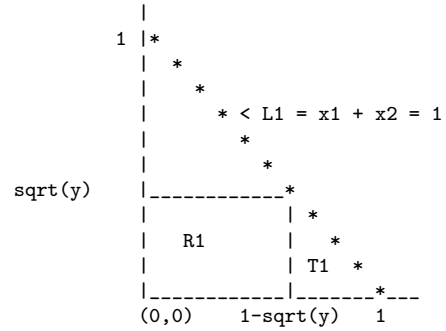
$$F_X(x)=P(X\leq x)=P(x_1^2\leq x)=P(x_1\leq\sqrt{x})=F_{X_1}(\sqrt{x})=\sqrt{x}.$$

Likewise, $F_Y(y)=\sqrt{y}$. Consider then

$$\begin{aligned}F_{X,Y}(x,y) &= P(X\leq x,Y\leq y)\\ &= P(x_1^2\leq x,x_2^2\leq y)\\ &= P(x_1\leq\sqrt{x},x_2\leq\sqrt{y})\\ &= P(x_1\leq\sqrt{x},x_2\leq\sqrt{y})\\ &= A(R_2)\\ &= \sqrt{xy}\\ &= F_X(x)F_Y(y).\end{aligned}$$



For the other part consider $F_{YZ}(y,1)$ and conclude that the r.v. Y and Z are not independent similarly. Then $F_{Y,Z}(y,1)$ is the area of R_1 and T_1.



■

9.23 Generating functions

Exercise 716.

Find the m.g.f of a discrete uniform random variable. Let X have $Z = \{1,2,\ldots,n\}$ with $f_X(x_i) = f_X(i) = P(X = i) = 1/n$.

Solution.

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= \sum_{i=1}^{n} e^{ti} P(X=i) \\ &= \sum_{i=1}^{n} e^{ti} \frac{1}{n} \\ &= \frac{e^t + e^{2t} + \ldots + e^{tn}}{n} \\ &= \frac{e^t(e^{nt}-1)}{n(e^t-1)}. \end{aligned}$$

If we use the expression

$$E(X) = m_X'(0) = \left(\frac{e^t + e^{2t} + \ldots + e^{tn}}{n}\right)' = \frac{1}{n}(e^0 + 1e^0 + 2e^0 + \ldots + ne^0) = (1+2+\ldots+n)/n = (n+1)/2.$$

$$E(X^2) = m_X''(0) = \left(\frac{e^t + e^{2t} + \ldots + e^{tn}}{n}\right)'' = \frac{1}{n}(e^0 + 1e^0 + 2e^0 + \ldots + ne^0) = (1+2^2+\ldots+n^2)/n = (n+1)(2n+1)/6.$$

Therefore $Var(X) = E(X^2) - E(X)^2 = \frac{n^2-1}{12}$. ■

Exercise 717.
(a) Find the probability generating function of the Binomial distribution i.e. $X \sim bin(n,p)$.
(b) Find the moment generating function of the Binomial distribution i.e. $X \sim bin(n,p)$.

Solution.
(a)

$$\begin{aligned} p_X(t) &= E(t^X) \\ &= \sum_{i=0}^{n} t^i P(X=i) \\ &= \sum_{i=0}^{n} t^i \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=0}^{n} \binom{n}{i} (tp)^i (1-p)^{n-i} \\ &= (tp+1-p)^n. \end{aligned}$$

(b) $m_X(t) = p_X(e^t) = (e^t p + 1 - p)^n$. ■

Exercise 718.
Find $p_X(0)$, $p'_X(0)$ for $X \sim bin(n,p)$.

Solution.
$P(X=0) = p_X(0) = (1-p)^n$.
$P(X=1) = p'_X(0) = np(1-p)^{n-1}$. ■

Exercise 719.
(a) Find the probability generating function of the Poisson distribution i.e. $X \sim Poisson(\lambda)$.
(b) Find the moment generating function of the Poisson distribution i.e. $X \sim Poisson(\lambda)$.

Solution.
(a)

$$\begin{aligned} p_X(t) &= E(t^X) \\ &= \sum_{i=0}^{n} t^i P(X=i) \\ &= \sum_{i=0}^{n} t^i e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \sum_{i=0}^{n} e^{-\lambda} \frac{(t\lambda)^i}{i!} \\ &= e^{-\lambda} \sum_{i=0}^{n} \frac{(t\lambda)^i}{i!} \\ &= e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}. \end{aligned}$$

(b) $m_X(t) = p_X(e^t) = e^{\lambda(e^t-1)}$. ■

Exercise 720.
Find $p_X(0)$, $p'_X(0)$, for $X \sim Poisson(\lambda)$.

Solution.
$P(X=0) = p_X(0) = e^{-\lambda}$.
$P(X=1) = p'_X(0) = \lambda e^{-\lambda}$. ■

Exercise 721.
(a) Find the variance of $X \sim E(\lambda)$.

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(b) Find the variance of $Y \sim N(\mu, \sigma^2)$.
Note that for $X \sim N(0,1)$, $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Then $F_X(x) = P(X \leq x)$. Moreover,

$$F_X'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Solution.
(a) $Var(X) = E(X^2) - E(X)^2 = E(X^2) - 1/\lambda^2$.

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \lambda \exp(-\lambda x) dx \\ &= -\int_0^\infty x^2 d \exp(-\lambda x) \\ &= +\int_0^\infty 2x \exp(-\lambda x) dx \\ &= \frac{2}{\lambda} \cdot \int_0^\infty x \lambda \exp(-\lambda x) dx \\ &= \frac{2}{\lambda} \cdot E(X) \\ &= \frac{2}{\lambda^2}. \end{aligned}$$

Therefore

$$Var(X) = E(X^2) - 1/\lambda^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

(b) $Var(Y) = E(Y^2) - 0 = E(Y^2)$. We have that for $X \sim N(0,1)$, the following.

$$\begin{aligned} E(X^2) &= \int_{-\infty}^\infty x^2 f_X(x) dx \\ &= \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= -\int_{-\infty}^\infty x d \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= (-x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}) \Big|_{-\infty}^\infty + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= (0) + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= (0) + 1 = 1. \end{aligned}$$

Therefore $Var(X) = E(X^2) - E(X)^2 = 1^2 - 0^2 = 1$. For $\sigma > 0$ we write $Y = \mu + \sigma X$. Then $Y \sim N(\mu, \sigma^2)$. Furthermore,

$$E(Y) = E(\mu + \sigma X) = \mu + \sigma E(X) = \mu,$$

and

$$Var(Y) = Var(\mu + \sigma X) = \sigma^2 Var(X) = \sigma^2.$$

■

Exercise 722.
What is the variance of X, $X \sim bin(n,p)$?

Solution.
$X = \sum_{i=1}^{n} Y_i$, where $Y_i \sim Bernoulli(p)$. Therefore

$$Var(X) = Var(\sum_{i=1}^{n} Y_i) = \sum_{i} Var(Y_i) = nVar(Y_i)$$

For a Bernoulli random variable $Y \sim Bernoulli(p)$ we have the following.

$$E(Y) = p \cdot 1 + (1-p) \cdot 0 = p, \quad E(Y^2) = p \cdot 1^2 + (1-p) \cdot 0^2 = p, \quad Var(Y) = E(X^2) - E(X)^2, \quad \Rightarrow Var(Y) = p - p^2 = pq.$$

Therefore, $Y_i \sim Y \sim Bernoulli(p)$ and we have the following.

$$Var(X) = nVar(Y_i) = npq = np(1-p).$$

■

Exercise 723.
Let $X \sim E(\lambda)$ so that

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(a) Find the m.g.f of X.

Solution.
$m_X(t) = E[e^{tX}]$. Consider the case $t < \lambda$ where $t - \lambda < 0$ and it will be used in the evaluation of $e^{(t-\lambda)x}$ for $x = \infty$.

$$\begin{aligned} E[e^{tX}] &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} dx \\ &= \frac{\lambda}{t-\lambda} \cdot \int_0^{\infty} de^{(t-\lambda)x} \\ &= \frac{\lambda}{t-\lambda} \cdot e^{(t-\lambda)x} \Big|_0^{\infty} \\ &= \frac{\lambda}{t-\lambda} \cdot (-1) = \frac{\lambda}{\lambda - t}. \end{aligned}$$

■

Exercise 724.

Let $X \sim E(\lambda)$; we have determined $m_X(t) = \frac{\lambda}{\lambda - t}$.
Let $X \sim bin(n,p)$; we have determined $m_X(t) = p_X(e^t) = (e^t p + 1 - p)^n$.
Let $X \sim Poisson(\lambda)$; we have determined $m_X(t) = p_X(e^t) = e^{\lambda(e^t - 1)}$.
Use $m_X(t)$ to determine $E(X), E(X^2), Var(X)$ for each of the three random variables.

Solution.
(a) We start with the exponential random variable. $m_X(0) = 1$ confirmed.

$$m_X'(0) = \frac{\lambda}{(\lambda - t)^2}(0) = 1/\lambda = E(X).$$

Moreover

$$m_X''(0) = \frac{2\lambda}{(\lambda - t)^3}(0) = 2/\lambda^2 = E(X^2).$$

We conclude then

$$Var(X) = 2/\lambda^2 - (1/\lambda)(1/\lambda) = 1/\lambda^2.$$

(b) We continue with the binomial random variable. $m_X(0) = 1$ confirmed.

$$m_X'(0) = n(e^t p + p - 1)^{n-1} \cdot p \cdot e^t(0) = np = E(X).$$

Moreover

$$m_X''(0) = pn(n-1)(e^t p + p - 1)^{n-2} p e^t e^t + np(e^t p + p - 1)^{n-1}(0) = n^2 p^2 - np^2 + np = E(X^2).$$

We conclude then

$$Var(X) = n^2 p^2 - np^2 + np - (np)^2 = np(1-p).$$

(c) We continue with the Poisson random variable. $m_X(0) = 1$ confirmed.

$$m_X'(0) = \frac{1}{e^\lambda} \cdot \lambda \ e^t e^{\lambda e^t}(0) = \lambda = E(X).$$

Moreover

$$m_X''(0) = \frac{\lambda}{e^\lambda}(e^t e^{\lambda e^t} + \lambda e^t e^t e^{\lambda e^t})(0) = \lambda(\lambda + 1) = E(X^2).$$

We conclude then

$$Var(X) = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

■

Exercise 725.
Find the m.g.f of the r.v. X that follows a geometric distribution $f_X(k) = q^{k-1}p$.

Solution.

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= \sum_{i=1}^{\infty} e^{ti} q^{i-1} p \\ &= \frac{p}{q} \sum_{i=1}^{\infty} e^{ti} q^i \\ &= \frac{p}{q} q e^t \frac{1}{1-qe^t} \\ &= \frac{pe^t}{1-qe^t}. \end{aligned}$$

Note that $qe^t < 1$ for convergence.. We can find $E(X)$ and $Var(X)$ then.

$$E(X) = m_X'(0) = \frac{pe^t}{(1-qe^t)^2}\Big|_{t=0} = \frac{p(1-q)+pq}{(1-q)^2} = \frac{1}{p}.$$

$$E(X^2) = m_X''(0) = \frac{pe^t + pqe^{2t}}{(1-qe^t)^3}\Big|_{t=0} = \frac{1+q}{(1-q)^2} = \frac{1+q}{p^2}.$$

Then

$$Var(X) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

■

Exercise 726.
Let a discrete r.v. X has m.g.f. $m_X(t)$. Find the m.g.f of $X+a$ and X/b.
Note that X,Y independent r.v.

$$m_{X+Y}(t) = m_X(t)m_Y(t).$$

Solution.
Let $Y = X+a$.

$$\begin{aligned} m_Y(t) &= E(e^{t(X+a)}) \\ &= e^{ta}E(e^{tX}) \\ &= e^{ta}m_X(t). \end{aligned}$$

Furthermore let $Z = X/b$.

$$\begin{aligned} m_Z(t) &= E(e^{tX/b}) \\ &= m_X(t/b). \end{aligned}$$

Thus if $X \sim N(0,1)$, let $Y = \frac{X-\mu}{\sigma}$. Then

$$m_Y(t) = e^{-\mu t/\sigma}m_X(t/\sigma).$$

■

Exercise 727.
Let X, Y follow a binomial distribution and are independent. Thus $X \sim bin(n,p)$ and $Y \sim bin(n,p)$. Then

$$f_X(k) = f_Y(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

Let $Z = X + Y$. Find the p.d.f of Z.

Solution.
We have $m_{X+Y}(t) = m_X(t)m_Y(t) = (pe^t + q)^n (pe^t + q)^n$. Moreover $p_{X+Y}(t) = (pt+q)^{2n}$. Then

$$p_{X+Y}(t) = (pt+q)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} (pt)^k q^{2n-k}$$

The coefficient of t^k is

$$p_{X+Y}(k) = \binom{2n}{k} p^k q^{2n-k}.$$

■

Exercise 728.
Let X_k be 1 if k-th toss is H and -1 if it is tails. Let $S_0 = 0$ and for $n > 0$ we define

$$S_n = X_1 + X_2 + \ldots X_n.$$

Solution.

■

Exercise 729.
Find the characteristic function of $X \sim Bernoulli(p)$.

Solution.

$$\begin{aligned} c_X(t) &= E(e^{itX}) \\ &= e^{it\cdot 1}\cdot p + e^{it\cdot 0}\cdot(1-p) \\ &= pe^{it} + (1-p) \\ &= p\cos t + ip\sin t + (1-p). \end{aligned}$$

■

9.24 Birthdays

Exercise 730.

Birthday problem.

In a room n people are gathered. What is the probability that (at least) two of them have the same birthday? Ignore leap years i.e. a year has 365 days and thus 365 birthdays are possible.
If E is the event that at least two people have the same birthday then E' is the event that no two (or even more) people share the same birthday.

Solution.
The first of the n people can have any of the 365 birthdays. Then the second one does not share a birthday with the first if the birthday is any one of the remaining 364 other than the first person's birthday. Thus the probability that the two have different birthdays is $364/365$. Continuing like this we generate the probability $P(E')$.

$$P(E') = \frac{365 \cdot 364 \cdot \ldots \cdot (365-(n-1))}{365^k}.$$

And thus

$$P(E) = 1 - P(E') = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365-(n-1))}{365^k}.$$

■

Exercise 731.

Birthday problem revisited: conditional probabilities.

In a room n people are gathered. What is the probability that (at least) two of them have the same birthday? Ignore leap years i.e. a year has 365 days and thus 365 birthdays are possible.

Solution.

Let A_2 be the event person 2 has different birthday from person 1. Let A_3 be the event person 3 has different birthday from persons 1,2. Let A_4 be the event person 4 has different birthday from persons 1,2,3. Etc. Let A_n be the event person n has different birthday from persons $1,2,\ldots,n$. Then

$$P(A_2) = 1 - 1/365.$$

Moreover,

$$P(A_3|A_2) = 1 - 2/365.$$

And by induction,

$$P(A_n|A_2A_3\ldots A_{n-1}) = 1 - (n-1)/365.$$

Then

$$p(E') = P(A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_n|A_2A_3\ldots A_{n-1})\ldots P(A_3|A_2)P(A_2) = (1-(n-1)/365)\ldots(1-2/365)(1-1/365).$$

We can then calculate $P(E) = 1 - p(E')$. ■

Exercise 732.

Birthday Problem generalized

We have m people that have birthdays that take n values, and let for siplicity they are drawn from the set $\{1,2,\ldots,n\}$. The probability that all m have different birthdays is

$$p = \frac{n(n-1)\ldots(n-m+1}{n^m}$$

How large should m be so that this p at least $1/2 = 1-1/2$, $1-1/4$, $1-1/8$, etc $1-1/2^{10}$.

Solution.

Let $m-1 = k\sqrt{n}$ where k is some small integer. We can also use that for $x \leq 1/2$ we have $e^{-2x} \leq 1-x$ or equivalently $1-x \geq e^{-2x}$. Then $(1-\frac{i}{n}) \geq e^{-2i/n}$.

$$\begin{aligned} p &= (1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})\ldots(1-\frac{m-1}{n}) \\ &\geq \exp(-2\cdot 1/n)\cdot\exp(-2\cdot 2/n)\cdot\exp(-2\cdot 3/n)\cdot\ldots\exp(-2\cdot(m-1)/n) \\ &\geq \exp(\frac{-2m(m-1)}{2n}) \geq \exp(-k^2) \end{aligned}$$

■

Exercise 733.

Planet CS has an n-day year. Among a population of k people of this planet find the expected number of triples of these people who have the same birthday. How large should k be for the expected value to be at least 1?

Solution.

The number of triples of people is $\binom{k}{3}$ and the persons in a triple have all the same birthday with probability $\frac{1}{n^2}$. Therefore the expected number of triples of people having the same birthday is $\binom{k}{3}\frac{1}{n^2}$. This value is 1 when $k \approx cn^{2/3}$ for a constant c. ■

9.25 Miscellanea

Exercise 734.

A coin is tossed with probability $p \in [0,1]$ coming H. If it is H, dealer pays player $1c$. If it is T, player pays dealer $2c$. What is the smaller p that player is not expected to lose money.

Solution.
$E(X) = p \cdot 1 + (1-p)(-2)$ is expected payoff. Since it is expected $E(X) \geq 0$ this is equivalent to $p \geq 2/3$. ■

Exercise 735.

Expected value. A random variable-free definition: An experiment has outcomes that take numerical values x_i. The probability that the outcome with value x_i appears is p_i. Then the expected value of the experiment is $E = \sum_i p_i x_i$.
(a) A (fair) coin is tossed four times. What is the expected number E of Heads?
(b) Number game. 100 numbers from 0 to 99 are to be used for drawing one number. A player can bet a dollar to a number. If the player wins the win is 67 dollars. Else player loses a dollar. What is the expected profit?

Solution.
(a) There are five outcomes when a coin is tossed four times: $0H, 1H, 2H, 3H, 4H$. The corresponding numerical values are $0, 1, 2, 3, 4$. It is easy to calculate that $E = 2$.
(b) With probability $1/100$ the profit is $66 = 67 - 1$. With $99/100$ the profit is -1 (a loss of 1).

$$99/100 \cdot (-1) + 1/100 \cdot 66 = -0.33$$

Thus if you play this game 1000 times you will lose on the average 330 dollars. ■

Exercise 736.

You flip a fair coin in a casino. You have 50% probability of winning. Let the bet be one dollar. You devise the following tactic: you decide to play repeatedly until the first way when you stop. Let N be the number of games (tosses) played.
(a) What is the expected number of games played i.e. $E(N)$?
(b) What is the fraction of Wins and thus what is the fraction of Losses in the N games played?
(c) What is the expected fraction of Wins and thus what is the expected fraction of Losses in the N games played?

Solution.
(a) The game is a geometric distribution with $1-p=p=0.5$ Thus the average number of games played before a win is $E(N)=1/p=2$.
(b) If N is the number of games, one is a win and thus $(N-1)$ is losses. The fraction of wins is $1/N$ and the factions of losses is $(N-1)/N=1-1/N$.
(c) If L is the losses $E(L)=E(N)-1=1$. We have 1 win out of N games. Thus the fraction of wins as an expected value.

$$\begin{aligned} E(1/N) &= \sum_{i=1}^{\infty} P(N=i)(1/i) \\ &= \sum_{i=1}^{\infty} (1-p)^{i-1} p(1/i) \\ &= \sum_{i=1}^{\infty} (1/2^i)(1/i). \end{aligned}$$

Using a previous sum manipulation using integration

$$\begin{aligned} \sum_{i=1}^{\infty} x^{i-1} &= \frac{1}{1-x} \Leftrightarrow \\ \sum_{i=1}^{\infty} \frac{1}{i} x^i &= -\ln(1-x) \Leftrightarrow \\ \sum_{i=1}^{\infty} \frac{1}{i} (1/2)^i &= -\ln(1/2) \Leftrightarrow \\ \sum_{i=1}^{\infty} \frac{1}{i} (1/2)^i &= \ln 2, \end{aligned}$$

we obtain the following result.

$$\begin{aligned} E(1/N) &= \sum_{i=1}^{\infty} (1/2^i)(1/i) \\ &= \ln 2 \approx 0.693. \end{aligned}$$

Note that the reciprocal of the expection $(1/2)$ is not equal to the expectation of the reciprocal of the random variable (0.693). For the expected fraction of losses.

$$E((N-1)/N) = E(1-1/N) = 1-E(1/N) = 1-\ln 2 \approx 0.307.$$

■

Exercise 737.

Alice-Bob game. There are two boxes containing dollar coins. One given to Alice the other to Bob. Both are told that one box contains X and the other $2X$. They are not told which box is which. Alice is given the possibility to switch.

Alice makes the following calculation. Let a be Alice's current take. With probability $1/2$ can end-up with $2a$ and with probability $1/2$ can end-up with $a/2$ if she switches with Bob. Thus the expected profit is

$$1/2(2a) + 1/2(a/2) = 5a/4 > a$$

Alice decides to switch.

Is Alice correct and rational?

What if Bob is given the opportunity to switch?

What if Alice or both are given ANOTHER opportunity to reconsider (and switch)?

What if this goes on ad infinitum?

Solution.

Alice is not correct. She makes the assumption that she has a, and then the possible outcomes are $a/2$ and $2a$. These are three different amounts of money. We know there are only two between the two of them, one is X the other is $2X$. Consider this example: the boxes have 10 and 20 dollars. If Alice has 10 dollars it can't be that Bob has 5 dollars. If Alice has 20 dollars it can't be that Bob has 40 dollars. In the first case the PROFIT for Alice is $20 - 10$ if she switches. In the latter case the PROFIT is $10 - 20 = -10$ in fact a loss if she switched. The expected profit is thus $10 - 10 = 0$. ■

Exercise 738.

St Petersburg's paradox. In a casino a (fair) coin is tossed. The number of tails T before the first head determines the payoff. Therefore is Z is the number of T before the first H, the payoff by the casino is 2^Z (say dollars). A player pays the casino once at entering the game.
(a) What is fair price for a player to pay the casino to enter the game? Explain.
(b) Now the casino changes its policy. The payof is fixed to $8,192$ dollars. Thus the award the casino would pay would never exceed $\min(10,000, 2^Z) = 2^{\min(13,Z)}$. What is fair price for a player to pay the casino to enter the game? Explain.

Solution.
(a) The probability of having Z tails and then a head is $(1/2^Z)(1/2) = 1/2^{Z+1}$. The payoff then is 2^Z. Therefore the expected payoff P is as follows.

$$E(P) = \sum_{Z=0}^{\infty} \frac{1}{2^{Z+1}} \cdot 2^Z = \sum_{Z=0}^{\infty} \frac{1}{2} = \infty.$$

Note that the number of T's before an H follows a geometric distribution with expected value $1/p$. Therefore on the average in two tosses we have a head. Thus the payoff is 2^1 as there is only one T on the average.
(b) The expected payoff now is as follows.

$$\begin{aligned} E(P) &= \sum_{Z=0}^{\infty} \frac{1}{2^{Z+1}} \cdot 2^{\min(13,Z)} \\ &= \sum_{Z=0}^{13} \frac{1}{2} + \sum_{Z=14}^{\infty} 2^{13} \cdot \frac{1}{2^{Z+1}} \\ &= 7 + 2^{13} \frac{1}{2^{14}} = 7 + 0.5 = 7.5. \end{aligned}$$

■

Exercise 739.

Linear Search. An array $A[1..n]$ contains n distinct and comparable keys. Part of the input is an arbitrary key value k. With probability p, key k is in A and it is equally likely to be in any of the n positions of A, and we want to find that i such that $A[i]$ is equal to k, but with probability $q = 1-p$ keys k is not in A and then lin search returns not FOUND (say $n+1$ value is not FOUND). What is the average number of comparisons performed in line 2?

```
  LinSearch(A[1..n],n,k)
1. for( i=1 ; i<=n ; i++ ) {
2.  if cmp(A[i],k)==EQUAL
3.    return(i); // FOUND at i
4. }
5. return(n+1);    // NOT FOUND
```

Solution.
If key k is in position i then i comparisons are to be performed. Then $x_i = i$. The probability that key is in i would be p/n. Furthermore, $x_{n+1} = n$ with $p_{n+1} = q = 1-p$. Then

$$\begin{aligned} E &= \sum_{i=1}^{n+1} p_i x_i = (p/n)\sum_{i=1}^{n} i + qn \\ &= \frac{pn(n+1)}{2n} + (1-p)n \\ &= \frac{p(n+1)}{2} + (1-p)n \\ &= \frac{(n+1)}{2} + (1-p)n \end{aligned}$$

If $q = 0$ and thus $1-p = 0$ and $p = 1$, then $E = (n+1)/2$. If $q = 1$ and thus $1-p = 1$ and $p = 0$, then $E = n$. If $0 < q < 1$, then $E < (n+1)/2 + n = (3n+1)/2$. ■

Exercise 740.

Binary search. Let N be the length of a sorted sequence $A[0..N-1]$, where $N = 2^n - 1$. We are looking for key k into A. What is, for Binary Search, (a) the minimal number of comparisons, (b) the maximal number of comparisons, (c) the average number of comparisons to find k (i.e. determine i such that $A[i] = k$) or determine k is not in A (and return say -1).

Solution.
(a) The minimal number of comparisons is obviously one.
(b) The maximal number of comparisons is n. In the worst case

$$C(2^n-1) = 1 + C((2^n-1-1)/2) = 1 + C(2^{n-1}-1) \Rightarrow c(n) = 1 + c(n-1), \text{ for } c(n) = C(2^n-1).$$

Solve for $c(n) = n = \lg(N+1)$.
(c) Calculate $P(X = i)$, where X is the number of comparisons performed to find an element of A. Let us split A into A_i, where A_1 is the set of keys of A that can be determined for equality with k in one comparison. Obviously $|A_1| = 1$, and the middle element $A[(N-1)/2]$ is the only element of A_1. Furthermore $|A_i| = 2^{i-1}$. Then

$$\begin{aligned} E(X) &= \sum_{i=1} nP(X=i)\cdot i \\ &= \sum_{i=1} n\frac{2^{i-1}}{2^n-1}\cdot i \\ &= \frac{1}{2^n-1}\cdot\sum_{i=1} n2^{i-1}\cdot i \\ &= \frac{1}{2^n-1}(n2^n - 2^n + 1) \\ &= (n-1) + \frac{n}{2^n-1} = (n-1) + \frac{n}{N}. \end{aligned}$$

We utilized

$$\begin{aligned} \sum_{i=0}^{n} x^i &= (x^{n+1}-1)/(x-1) \Rightarrow \\ \sum_{i=0}^{n} i\cdot x^{i-1} &= (n+1)x^n(x-1) - (x^{n+1}-1)/(x-1)^2 \Rightarrow \\ \sum_{i=0}^{n} i\cdot 2^{i-1} &= 2^n\cdot n - 2^n + 1. \end{aligned}$$

Note that the analysis applies to the case where k is IN A. If k is not in A, then n comparisons are needed to draw that conclusion. ■

Exercise 741.
Let a sample space $S = \{1,2,3,\ldots\}$. Let $P(n) = 1/2^n$ for all $n = 1,2,3,\ldots$.
What is the probability that an outcome (pick a positive integer) is an even number?

Solution.

$$P(even) = 1/4 + 1/16 + \ldots = \sum_{i=1}^{n} \frac{1}{2^{2i}} = \sum_{i=1}^{n} \frac{1}{4^i} = 1/4(1 + 1/4 + 1/4^2 + \ldots) = 1/4(1/(1-1/4)) = (1/4)(4/3) = 1/3.$$

■

Exercise 742.

You are given a random vector $a = (a_1, \ldots a_n)$, where a_i is equally likely and independently to be 0 or 1, i.e. $P(a_i = 1) = P(a_i = 0) = 1/2$. Answer the following questions.
(a) **(Warmup)** What's the probability that a is the all zero vector ?
(b) Suppose that a, b are two 0-1 vectors of length n whose components were chosen uniformly at random as discussed previously. What is the expected value of the inner product $a \cdot b = \sum_{i=1}^{n} a_i b_i$? Explain.
(c) Let d be a vector of integers mod p (i.e. elements of d are $0, \ldots p-1$), where p is a prime. Let a be a random vector of 0-1's chosen as before. What is an upper bound on the probability that $\sum d_i a_i \equiv 0 \bmod p$? Explain.

Solution.
(a) $1/2^n$.
(b) $n/4$.

$$E[c] = E[ab] = E[\sum_i a_i b_i] = \sum_i E[a_i b_i]$$

$a_i b_i$ is 0 with probability $3/4$ and 1 with probability $1/4$ (when both $a_i = b_i = 1$).

$$E[a_i b_i] = 0(3/4) + 1(1/4) = 1/4$$

Therefore

$$E[c] = E[ab] = E[\sum_i a_i b_i] = \sum_i E[a_i b_i] = n(1/4)$$

(c) The vector d is given (and is not necessarily random). Assume $d \neq 0$, i.e. at least one component of the vector is non-zero since otherwise the problem is trivial. $\sum d_i a_i \equiv 0 \bmod p$.

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \bmod p \\ d_1 a_1 + d_2 a_2 + \ldots + d_n a_n &\equiv 0 \bmod p \\ d_1 a_1 &\equiv (-d_2 a_2 - \ldots - d_n a_n) \bmod p \\ d_1 a_1 &\equiv Z \bmod p \end{aligned}$$

where $Z \equiv (-d_2 a_2 - \ldots - d_n a_n) \bmod p$. Then,

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \bmod p \\ d_1 a_1 &\equiv Z \bmod p \\ a_1 &\equiv (d_1)^{-1} Z \bmod p \\ a_1 &\equiv B \bmod p \end{aligned}$$

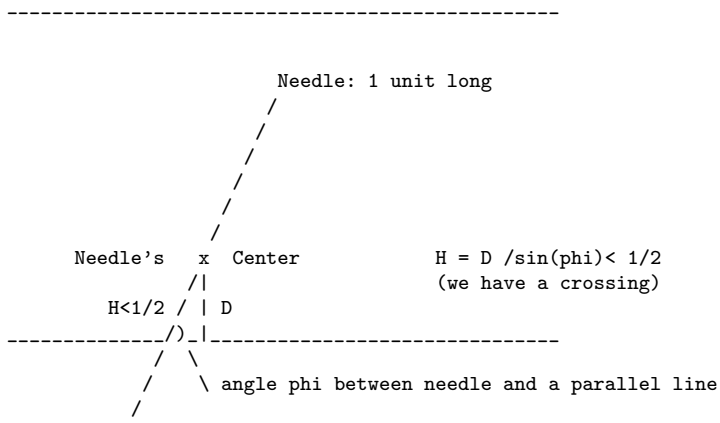
The inverse of d_1 exists since $d_1 x \equiv 1 \bmod p$ has a single solution for x by the fact that $d_1 < p$ is such that $(d_1, p) = 1$ and p is prime. $B = (d_1)^{-1} Z \bmod p$ is an integer in $0, \ldots, p-1$. a_1 is a (uniformly at) random (chosen) 0,1. What is the probability that the random a_1 is B? Naturally this probability is at most $1/2$, as after we fix B, a_1 can agree with this fixed value of B half of the time only. If the a_i's were not binary but ternary, then the probability bound would be $1/3$ instead. ■

Exercise 743.

Needles. We draw parallel lines one unit apart. A needle of length one unit is thrown at random. We observe whether the needle crosses one of the parallel lines. What is the probability of this happening

Hint. Consider the distance D from the middle of the needle to the closest parallel line. Consider also the angle of the needle with the parallel lines.

Solution.



If ϕ is the angle between the needle and the parallel lines (consider $\phi \leq 90$ degrees). Moreover $0 \leq D \leq 1/2$, and $0 \leq \phi \leq \pi/2$. A crossing occurs when $D/\sin(\phi) < 1/2$ as shown above. For a throw of the needle D and ϕ are uniformly at random drawn from the ranges indicated.

$$f_D(d) = \begin{cases} 2 & 0 \leq D \leq 1/2 \\ 0 & \text{otherwise,} \end{cases}$$

$$f_\Phi(\phi) = \begin{cases} 2/\pi & 0 \leq \phi \leq \pi/2 \\ 0 & \text{otherwise,} \end{cases}$$

Then $f_{D,\phi} = \frac{4}{\pi}$ for $0 \leq D \leq 1/2$ and $0 \leq \phi \leq \pi/2$ and 0 otherwise. For a crossing to occur $D/\sin(\phi) < 1/2$, or equivalently, for the probability density function of the center of the needle distance D from a parallel line we have The needle crosses a line for $D < 1/2\sin(\phi)$. The probability that the needle crosses a parallel line is then

$$P = \int_0^{\pi/2} \int_0^{1/2\sin\phi} \frac{4}{\pi} dD d\phi = 2/\pi.$$

■

Exercise 744.

Monte-Carlo approximation of π or not.

(a) We throw darts on or inside a unit square. Insider the unit square there a circle inscribed with diameter one. If darts are thrown uniforlmy at random and all hit the square, what is the probability that they hit the circle (i.e. fall on or insider the inscribed circle)?

(b) What is the probability that a dart hits a circular UNIT radius target at distance x, where $0 \leq x \leq 1$.

Solution.

The probability a dart hits the circle is the area of the circle over the area of the square. Square S has area

$$a(S) = 1 \cdot 1 = 1,$$

and the circle with diameter 1 has radius $1/2$ and thus area

$$a(C) = \pi \cdot (1/2)^2 = \pi/4.$$

Therefore if E is the event the dart hits the circle,

$$p(E) = a(C)/a(S) = \pi/4$$

In a montercarlo simulation we throw N darts and count the ones hitting inside the circle $n \leq N$. The by the law of large numbers $n/N \approx \pi/4$ and thus an approximation to π would be

$$\text{approx}(\pi) \approx \frac{4n}{N}.$$

(b) At distance x the area of the circlue is πx^2 versus $\pi \cdot 1^2$ the area of the target circle. Thus the probability of hitting the target within distance x is x^2. ■

9.26 Binomial process and Bernoulli trials

Exercise 745.

Examine the monotonicity of the binomial process $B(n,p)$.

Solution.
Let $B(n,p;k)=\binom{n}{k}p^k(1-p)^{n-k}$. Consider

$$\frac{B(n,p;k)}{B(n,p;k-1)}=1+\frac{(n+1)p-k}{k(1-p)}.$$

If $k<(n+1)p$ the binomial terms form an increasing sequence. If $k>(n+1)p$ the binomial terms form a decreasing sequence. If $(n+1)p=t$ then $B(n,p;k)=B(n,p;k-1)$, since the fraction above is 1. Otherwise there exists only one integer m such that $(n+1)p-1<m\leq(n+1)p$. ■

Exercise 746.

Let $S_n = X_1 + \ldots + X_n$, where X_i are individually independent Bernoulli processes with $X_i \sim b(p)$. Then $S_n \sim B(n,p)$. Given the expectation, and the variance of S_n and show that

$$P(S_n \geq t) \leq t(1-p)/(t-np)^2$$

for $t > np$.

Solution.
if $t > np$ this implies $t \geq np+1 \geq (n+1)p$ since $p \leq 1$. Let $B(n,p;k) = \binom{n}{k} p^k (1-p)^{n-k}$. Consider

$$\frac{B(n,p;k)}{B(n,p;k-1)} = 1 + \frac{(n+1)p-k}{k(1-p)}.$$

If $k < (n+1)p$ the binomial terms form an increasing sequence. If $k > (n+1)p$ the binomial terms form a decreasing sequence. If $(n+1)p = t$ then $B(n,p;k) = B(n,p;k-1)$, since the fraction above is 1. Otherwise there exists only one integer m such that $(n+1)p - 1 < m \leq (n+1)p$. ■

9.27 Chernoff and Hoeffding bounds

Exercise 747.
Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(p)$. Then $S_n \sim B(n,p)$. Show that for $6np < n$ we have the following

$$P(S_n \geq 6np) \leq 2^{-6pn},$$

using Chernoff bounds.

Solution.
For $\delta > 2e-1$ in a Chernoff bound, the following applies.

$$P(S_n \geq (1+\delta)pn) \leq 2^{-(1+\delta)pn}.$$

If we pick $(1+\delta) = 6$ the result follows. ■

Exercise 748.

Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(1/2)$. Then $S_n \sim B(n, 1/2)$. In n coin tosses ($p = 1/2$ of a fair coin), what is the probability of having r Heads, where $r < n/2$?

$$P(r\ Heads,\ r < n/2)?$$

Use a Chernoff or Hoeffding bound. Do not browse later questions.
Use the bound to answer the following question: What is the probability that in 100 coin tosses we have 25 or fewer heads?

Solution.
Use a chernoff bound

$$P(S_n \leq (1-\delta)np) \leq \exp(-\delta^2 np/2).$$

with $(1-\delta)n/2 = m$ i.e. $\delta = 1 - 2m/n$. Now $n = 100$, $m = 25$, $\delta = 1 - 50/100 = 1/2$ we have

$$P(S_n \leq 25) \leq \exp -(1/4)25 < 2/1000.$$

■

Exercise 749.

(Coin tossing continues.) Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(1/2)$. Then $S_n \sim B(n, 1/2)$.
In n coin tosses ($p = 1/2$ of a fair coin), what is the probability of having r Heads, where $r < n/2$?

$$P(|S_n/n - 1/2| \geq \delta)$$

Use a Chebyshev and then a Chernoff or Hoeffding bound.
Consider the cases $\delta = 1/2$, $\delta = \sqrt{1.5\ln(n)/n}$ and $\delta = 2\sqrt{1.5\ln(n)/n}$. Observe variation when delta doubles for the last two choices.

Solution.
Chebyshev first.

$$P(|S_n/n - 1/2| \geq \delta) = P(|S_n - n/2| \geq \delta n) \leq \frac{Var(S_n)}{\delta^2 n^2} \leq \frac{1}{4\delta^2 n}.$$

Note that $Var(S_n) = npq = n(1/2)(1/2) = n/4$.
The term below (n/2) is introduced to form $np = n/2$ next.

$$P(|S_n/n - 1/2| \geq \delta) = P(|S_n - n/2| \geq \delta n) = P(|S_n - n/2| \geq 2\delta(n/2)) \leq 2\exp(-(n/2)4\delta^2/3) \leq 2\exp(-(2/3)n\delta^2).$$

The latter one
(a) for $\delta = 1/2$ becomes $2e^{-n/6}$,
(b) for $\delta = \sqrt{1.5\ln(n)/n}$ becomes $2/n$, and
(c) for $\delta = 2\sqrt{1.5\ln(n)/n}$ becomes $2/n^4$.
Consider now a Hoeffding bound.

$$P(|S_n/n - 1/2| \geq \delta) = P(|S_n - n/2| \geq 2\delta(n/2)) \leq 2\exp(-2n\delta^2).$$

The latter one
(a) for $\delta = 1/2$ becomes $2e^{-n/2}$,
(b) for $\delta = \sqrt{1.5\ln(n)/n}$ becomes $2/n^3$, and
(c) for $\delta = 2\sqrt{1.5\ln(n)/n}$ becomes $2/n^{12}$. ■

Exercise 750.

Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(p)$. Then $S_n \sim B(n,p)$. Show that for $r > np$,

$$P(S_n - np \geq r) \leq \left(\frac{npe}{r}\right)^r.$$

Solution.
We use the Chernoff trick with the moment generating function. For this we note the following.

$$E(\exp t(X_i - p)) = p \cdot e^{t(1-p)} + (1-p) \cdot e^0 = pe^{t(1-p)} + 1 - p \leq pe^t + 1 \leq e^{pe^t}.$$

In the application of the Chernoff trick, the first inequality is by Markov's inequality's application. The last step uses the result above

$$\begin{aligned} P(S_n - np \geq r) &= P(t(S_n - np) \geq tr) \\ &= P(\exp t(S_n - np) \geq \exp tr) \\ &\leq e^{-tr} E(\exp(t(S_n - np))) \\ &= e^{-tr} E(\exp(t(\sum_i X_i - np))) \\ &= e^{-tr} \prod_i E(\exp(t(X_i - p))) \\ &= e^{-tr} \prod_i e^{(pe^t)}) \end{aligned}$$

We have thus concluded the following.

$$\begin{aligned} P(S_n - np \geq r) &\leq e^{-tr} \prod_i e^{pe^t} \\ &= e^{-tr} e^{npe^t} \\ &= e^{-tr + npe^t} \end{aligned}$$

The exponent is $f(t) = -tr + npe^t$. We find $f'(t) = -r + npe^t$. We then equate it to zero and solve for t that is, $f'(t) = -r + npe^t = 0$ implies $e^t = r/(np)$ and consequently $t = \ln(r/(np))$. Deriving $f''(t) = npe^t = np(r/(np)) = r > 0$. Thus $f(t)$ has a minimum at $t = \ln(r/(np))$. Substituting for t the right-hand side value we have.

$$e^{-\ln(r/(np))r + np(r/(np))}) = (\frac{np}{r})^r e^r = (\frac{npe}{r})^r.$$

■

Exercise 751.
Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(p)$. Then $S_n \sim B(n,p)$. Show the following for $1 \leq r < n$.

$$P(S_n \geq r) \leq \left(\frac{npe}{r}\right)^r.$$

$$P(S_n < r) \leq \binom{n}{n-r}(1-p)^{n-r}.$$

Solution.
We prove a variation of the previous problem i.e.

$$P(S_n \geq r) \leq \left(\frac{npe}{r}\right)^r.$$

Consider all subsets S of cardinality r of $\{1,2,\ldots,n\}$. There are $\binom{n}{r}$ of them and for $j \in S$ we consider experiment X_j a success. We call the event related to the given S, event E_S (where all experiments described by S are a success). $P(E_S) = p^r$. Thus $P(S_n \geq r)$ is upper bounded by the union of all such events possible.

$$P(S_n \geq r) \leq P(\cup E_S) \leq \sum_{S,|S|=r} P(E_S) \leq \binom{n}{r} p^r. \tag{9.2}$$

However, the problem in question asks for a bound of the following.

$$P(S_n < r)$$

The probability of having fewer than r successes is also the probability of having at least $n-r$ failures. This is Eq.(9.2) substituting failure for success and thus probability of failure $1-p$ for probability of success p. Therefore

$$P(S_n < r) = P(B(n,p;k) < r)P(B(n,(1-p);k \geq n-r) \leq \binom{n}{n-r}(1-p)^{n-r}.$$

■

Exercise 752.

Let $S_n = \sum_{i=1}^n X_i$, where X_i are individually independent Bernoulli processes with $X_i \sim b(p)$. Then $S_n \sim B(n,p)$. Show that for $r \leq np$,

$$P(S_n < r) \leq \frac{r(1-p)}{np-r} B(n,p;r).$$

Solution.
Consider

$$P(S_n < r) = P(B(n,p) < r) = \sum_{i=0}^{r-1} B(n,p;i).$$

Now take the ration

$$\begin{aligned} \frac{B(n,p;i-1)}{B(n,p;i} &= \frac{i}{n-i+1}\frac{1-p}{p} \\ &\leq \frac{r}{n-r}\frac{1-p}{p} \\ &= \frac{rq}{(n-r)p}. \end{aligned}$$

The inequality above follows from $1/(n-i+1) < 1/(n-r) \Leftrightarrow r > i$. We then call a the ratio $a = rq/((n-r)p)$. Note that $a \leq 1$ since $r(1-p) \leq (n-r)p \Leftrightarrow r \leq pn$, with the latter given as a condition in the problem statement. Therefore for all $i = 0, \ldots, r-1$

$$B(n,p;i-1) \leq aB(n,p;i),$$

which implies

$$\begin{aligned} \sum_{i=0}^{r-1} B(n,p;i) &\leq \sum_{i=1}^{r} a^r B(n,p;r) \\ &\leq B(n,p;r)\sum_{i=1}^{r} a^r B(n,p;r) \\ &\leq B(n,p;r)\sum_{i=1}^{\infty} a^r B(n,p;r) \\ &\leq \left(\frac{1}{1-a} - 1\right) B(n,p;r). \\ &\leq \frac{a}{1-a} B(n,p;r). \end{aligned}$$

Note that $a/(1-a) = rq/(np-r)$. ■

9.28 Permutations and Derangements

Exercise 753.

Give the number of derangements of permutations on n elements $\{1,2,\ldots,n\}$.
A permutation π is a derangement if and only if it does not have a fixed point. Point i is a fixed point of a permutation π if $\pi(i)=i$. Therefore in a derangement π we have $\pi(i)\neq i \quad \forall i,\ i=1,\ldots,n$.

Solution. Let S be the set of all permutations. Let $A_i=\{\pi\in S:\pi(i)=i\}$. A permutation π is a derangement if and only if

$$\pi\in\overline{A_1\cup\ldots\cup A_n}$$

Moreover A_I, $I\subseteq\{1,2,\ldots,n\}$ with $|I|=k$ contains all permutations with k fixed points. There are $|A_I|=(n-k)!$ of them. Let us try to count instead

$$n!-A_1\cup\ldots\cup A_n$$

or

$$A_1\cup\ldots\cup A_n$$

By the principle of inclusion exclusion we get

$$|A_1\cup\ldots\cup A_n|\sum_i|A_i|-\sum i<j|A_i\cap A_j|+\sum i<j<k|A_i\cap A_j\cap A_k|-\ldots=\sum_{i=1}^{n}(-1)^{i+1}\binom{n}{i}(n-i)!=n!\sum_{i=1}^{n}(-1)^{i+1}/i!.$$

Thefore $n!-|A_1\cup\ldots\cup A_n|=\sum_{i=0}^{n}(-1)^{i+1}/i!$. ■

Exercise 754.

Give the number of derangements of permutations on n elements $\{1,2,\ldots,n\}$.

Solution.
An inductive approach. Let $p(n)$ be the number of derangements of n objects. Let $p(1)=j$. If $p(j)=1$ then we are left with $n-2$ objects and positions. The are $n-1$ choices for j and $p(n-2)$ choices for derangements on $n-2$ objects. If on the other hand $p(1)=j$ but $p(j)\neq 1$, the for the $2,\ldots,n$ objects we have $1,2,\ldots,j-1,j+1,\ldots,n$ i.e. $n-1$ numbers to derange. Therefore combining the two cases

$$p(n)=(n-1)p(n-2)+(n-1)p(n-1).$$

Furthermore, by induction

$$p(n)=np(n-1)+(-1)^n,$$

as follows.

$$\begin{aligned}
p(n) &= (n-1)p(n-2)+(n-1)p(n-1)\\
&= (n-1)p(n-1)+(n-1)p(n-2)\\
&= (n-1)p(n-1)+p(n-1)-p(n-1)+(n-1)p(n-2)\\
&= np(n-1)-p(n-1)+(n-1)p(n-2)\\
&= np(n-1)-(n-1)p(n-2)-(-1)^{n-1}+(n-1)p(n-2)\\
&= np(n-1)-(-1)^{n-1}\\
&= np(n-1)+(-1)^n.
\end{aligned}$$

Then $q(n)=p(n)/n!$ gives the following.

$$\begin{aligned}
p(n) &= np(n-1)+(-1)^n \Leftrightarrow\\
q(n) &= n\frac{p(n-1)}{n!}+(-1)^n/n!\\
q(n) &= q(n-1)+(-1)^n/n!.
\end{aligned}$$

Summing the last equation for $n=2$ to n we obtains

$$q(n)=\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}.$$

Then

$$p(n)=n!\left(\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}\right).$$

■

Exercise 755.

(a) Calculate $P(n,k)$ the number of permutations of n objects taken k at a time.
(b) Calculate $C(n,k)$ the number of combinations of n objects taken k at a time. Order in the set (of a combination) makes no difference.

Solution.
(a) We have n possibilities for the first element of the permutation, $n-1$ for the second, etc. Total is

$$P(n,k) = n(n-1)\dots(n-(k-1)) = \frac{n!}{(n-k)!}$$

(b) We have $P(n,k)$ possibilities but then need to adjust because order is not important by dividing with $k!$. Total is

$$C(n,k) = P(n,k)/k! = n(n-1)\dots(n-(k-1)/k! = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

■

Exercise 756.
Let Y be the number of fixed points in a random permutation of n distinct objects. Show that $E(Y) = 1$.
Example. Consider two objects $\{1,2\}$ One permutation is $p(1) = 1, p(2) = 2$ with 2 fixed points; the other one is $q(1) = 2, q(2) = 1$ with 0 fixed points. THe expected number is 1!

Solution.
Let X be a random permutation. Let X_i be 1 if X fixes i and 0 otherwise. Then let Z denote the number of fixed points.

$$Z = X_1 + X_2 + \ldots + X_n \Rightarrow E(Z) = E(X_1 + X_2 + \ldots + X_n) = \sum_E (X_i) = nE(X_i) = n(1/n) = 1.$$

We used $E(X_i) = 1/n$. ■

Exercise 757.
A coin is tossed once. Let X be a r.v. with $X = 1$ for Heads and $X = 0$ for Tails. Let $Y = 1 - X$. So that X and Y are not independet.

Solution.
We have $E(X) = 1/2 = E(Y)$. Moreover $X \cdot Y = 0$.

$$0 = E(X \cdot Y) \neq E(X) \cdot E(Y) = 1/2 \cdot 1/2.$$

■

Exercise 758.
Let (x,y) be a point of a unit cicle chosen uniformly at random. We have seen before for $R=\sqrt{x^2+y^2}$, $F_R(r)=r^2$ for $0\leq r\leq 1$. Moreover $f_R(r)=2r$. What is the expected value $E(R)$?

Solution.

$$E(R)=\int_0^1 rf_R(r)dr=\int_0^1 r2rdr=\int_0^1 d2r^3/3==2/3$$

■

9.29 Selections and Bin and Balls

Exercise 759.
What is the probability that among three digits integers with all distinct digits we pick a three digit denary integer that has no 0 (leading or otherwise), 8 and 9 and all of its digit are distinct.

Solution.
Viewing integers as ordered this means we have 7 choices for the first digit, 6 for the second and 5 for the third for a total of 210. Our universe of three digits integers with distinct digits is $10 \cdot 9 \cdot 8 = 720$. Thus the probability of generating one is $210/720$. ■

Exercise 760.

(a) In an urn we have 50 balls of which 25 are red and 25 are blue. We select 10 balls from the urn. What is the probability that 5 of the balls are red and 5 are blue (of the 10 drawn)? Assume balls are selected with replacement.
(b) Repeat without replacement.

Solution.
(a) There are $|S| = 50^{10}$ possible selections of 10 balls with replacement. There are $25^5 \cdot 25^5$ ways to pick 5 red balls, followed by 5 blue balls. But this is one possible scenario out of the $\binom{10}{5}$ possible ones. So for the event A of having 5 red balls and 5 blue balls we have that $|A| = 25^5 \cdot 25^5 \cdot \binom{10}{5}$. Therefore the probability $P(A)$ for event A is calculated as

$$P(A) = \frac{25^5 \cdot 25^5 \cdot \binom{10}{5}}{50^{10}} = \frac{252}{1024} \approx 0.25.$$

(b) What if the selection is without replacement? Then $|S| = 50!/40! = 50 \cdot 49 \cdot \ldots 41$. Then, $|A| = (25!/20!)^2 \cdot \binom{10}{5}$. Then

$$P(A) = \frac{25! \cdot 25! \cdot 40! \cdot \binom{10}{5}}{20! \cdot 20! \cdot 40!} \approx \frac{25 \cdot 2 \cdot 252}{20 \cdot 1024 \cdot \sqrt{3}} \approx 0.35,$$

where we used Stirling's approximation for the factorial for $20!, 25!, 40!$. ■

Exercise 761.

Find the number of ways you can throw n balls into N bins for the following scenarios.
(a) Distinct balls into distinct bins.
(b) Indistinct balls into distinct bins.
(c) Indistinct balls into distinct bins with at least one ball per bin.
(d) Indistinct balls into distinct bins with at least k_i balls into bin i?
(e) Indistinct balls into distinct bins at most one ball per bin.
(f) Distinct balls into distinct bins at most one ball per bin.
(g) Distinct balls into indistinc bins at least one ball per bin.
(h) Distinct balls into indistict bins at most one ball per bin.
(i) Indistinct balls into indistict bins at most one ball per bin.

Solution.
(a) We describe the distribution as a n-digit sequence, each digit taking one of N values representing a bin's ID (e.g. from 1 to N). There are

$$\underbrace{N \times N \times \ldots \times \ldots N}_{\text{since there are } \ n \ \text{ digits}} = N^n$$

(b) This is equivalent to n_i the number of balls in bin i: $\sum_{i=1}^{N} n_i = n$. In other words, the number of solutions n_i of this equation in the non-negative integers is the answer to the original question. This is $\binom{n+N-1}{N-1} = \binom{n+N-1}{n}$. Think also $N+1$ vertical lines (inclusive of first and last). First and last are fixed, they can be ignored thus the position of the remaining $N-1$ varies. Total number is out of $n+N-1$ marked positions we pick $N-1$ and turn them into vertical line markers $\binom{N+n-1}{N-1}$. Between the markers we count balls per corresponding bin.
(c) Place one indistinct ball per bin. The remaining $n-N$ balls remain to be distributed into the N bins. By case (b) we have $\binom{n-1}{N-1}$. Equivalently, the n balls create $n-1$ gaps between them. Pick $k-1$ out of the $n-1$ gaps in $\binom{n-1}{N-1}$.
(d) Similar to (c). Put k_i balls as neede into bin i. The remaining balls are handled as in (b) above.

$$\binom{n-\sum_i k_i}{N-1}.$$

(e) Implied is that $N \geq n$. We need to choose which n of the N urns will be assigned a ball! Thus $\binom{N}{n}$.
(f) N choice for first ball, $N-1$ for second and so on. Total $N(N-1)\ldots(N-n+1)$. Note that if $N=n$, then the answer is $n!$. Moreover, $n > N$, answer is 0.
(g) That's a Stirling number $\left\{ \begin{matrix} N \\ n \end{matrix} \right\}$.
(h) Not many choices: if $n < N$ the answer is 1; otherwise it is zero!
(i) See the previous question. Not many choices: if $n < N$ the answer is 1; otherwise it is zero! ■

9.29.1 ... or not

Exercise 762.

We throw n (indistinguishable) balls into N (indistinguishable) bins. Label them $1,\ldots,n$ and $1,\ldots,N$ respectively.
(a) What is the probability all balls fall into bin k? What is the probability that bin k is empty?
(b) What is the expected number of balls in bin k?
(c) What is the expected number of empty bins ?

Solution.
(a) All balls fall into bin k with probability $(1/N)^n$. Bin k remains empty with probability $(1-1/N)^n$.
(b) Let us use a random variable X_i is 1 if ball i goes to bin k, and 0 if it goes to another bin. Then $Y_k = \sum_{i=1} nX_i$ counts the number of balls into bin k. We have

$$E(X_i) = 1 \cdot 1/N + 0 \cdot (1-1/N) = 1/N.$$

Then using the linearity of expectation.

$$E(Y_k) = E(\sum_{i=1}^{n} X_i) = \sum_i E(X_i) = n \cdot 1/N = n/N.$$

(c) Let A_k be the event that bin k is empty. Then, $p(A_k)$ is the probability bin k is empty. Let B_k be a random variable that is 1 if bin k is empty and 0 otherwise. Then, $B = \sum_k B_k$ counts the number of empty bins. First,

$$P(A_k) = (1-1/N)^n$$

Then by linearity of expectation.

$$E(B) = E(\sum_k B_k) = \sum_k E(B_k)$$

Finally

$$E(B_k) = 1 \cdot P(A_k) + 0 \cdot (1 - P(A_k) = (1-1/N)^n$$

We conclude that

$$E(B) = \sum_k E(B_k) = N(1-1/N)^n.$$

The latter is $\approx Ne^{-n/N}$.
If $n = N$, it says that $1/e$ of the N bins are empty. If we throw n more balls $1/e \cdot 1/e \cdot N$ of the bins will be empty. How many times m do we need to repeat this so that $(1/e)^m N < 1$? Solving for m we find $m > \ln(N)$.
Let C be the event that there is an empty bin. Then we have the following.

$$P(C) = P(\exists eptybin) = P(\cup_{k=1}^{N} empty - bink) \leq \sum_k P(empty - bink) \leq N(1-1/N)^n.$$

We could have proved the same by Markov's inequality.

$$P(B \geq 1) \leq E(B)/1 = N(1-1/N)^n.$$

What is n so that $P(C)$ is w.h.p? Let us say we want $P(C) \leq 1/N$ or equivalently we want $Ne^{-n/N} \leq 1/n$. Set $n = 2N\ln N$. Then $Ne^{-n/N} \leq 1/N$ indeed. ■

Exercise 763.

We throw n (indistinguishable) balls into N (indistinguishable) bins. Label them $1,\ldots,n$ and $1,\ldots,N$ respectively. Continuing the previous problem for what value of n do we expect to see two balls in one bin, say bin k? This assumes there is one ball already into bin k.

Solution. The probability p of no second ball is as follows. We use $e^x \geq 1+x$ for all $x \in R$ in the form $e^{-x} \geq 1-x$.

$$1 \cdot (1-\frac{1}{N}) \cdot (1-\frac{2}{N}) \cdot (1-\frac{3}{N}) \ldots \cdot (1-\frac{n-1}{N}) \leq e^{-1/N} e^{-2/N} \ldots e^{-(n-1)/N} = e^{-n(n-1)/(2N)}$$

The birthday paradox says that if we bound this probability by $1/2$

$$e^{-n(n-1)/(2N)} \leq 1/2$$

the probability of a collision becomes $\geq 1/2$. Solving the equation above for n we have $n \approx \sqrt{2\ln(2)N}$. For a planet (Earth) with 365 days on a non-leap year this means that in a group of 23-24 people there would be two with same birthday (with probability at least 0.5). ■

Exercise 764.

Assume $n = N$. What is the probability that bin k has at least t balls in it?

Solution. Our current scenario simplifies to n balls into n bins the fullest bin has $O(\lg n \ \lg\lg n)$ balls w.h.p.

Let bin k has t balls. The probability this is the case is $B(n,p;t)$. A reminder

$$B(n,p;t) = \binom{n}{t} p^t (1-p)^{n-t}.$$

The following result in the form of Equations (9.4) and (9.5) appears in Feller [18] and in its complete form including Equation (9.3) as Theorem 1.1 in Bóllobás [4], where $B(n,p;m)$ is expanded and bounded,

Theorem 9.41 **Feller[18], Bóllobás[4]**

Let random variables X_i be inpdependent and follow a Bernoulli distribution that is, $X_i \sim b(p)$, and let $S_n = \sum_{i=1}^{n} X_i$. Then for $m = \lceil upn \rceil$, $u > 1$,

$$P(S_n \geq m) \leq \frac{u}{u-1} B(n,p;m). \tag{9.3}$$

Furthermore,

$$P(S_n \geq m) \leq \frac{m(1-p)}{(m-np)^2}. \tag{9.4}$$

and

$$P(S_n \leq m) \leq \frac{(n-m)p}{(np-m)^2}. \tag{9.5}$$

The first form appears in [4]; the other two in [18].

By Eq.(9.3) with $t = \lceil upn \rceil$, $p = 1/n$ and $t = \lceil u \rceil$, $u > 1$. In Eq.(9.3) the $u/(u-1)$ becomes at most 1.5 for $u > 3$ or equivalently $t > 3$. Therefore for $t > 3$

$$P(Bin\ k \geq t) = P(S_n \geq t) \leq \frac{u}{u-1} B(n,p;t) \leq 1.5 \binom{n}{t} (1/n)^t. \tag{9.6}$$

This bound makes sense if we can put $(1-p)^{n-t}$ into use that we obviously did not do.

An alternative is to argue as follows (for bin k that is).

$$P(Bin\ k \geq t) \leq \sum_{S_n \subseteq \{1,\ldots,n\}, |S_n| = t} \prod_i P(\text{ball i } \in S_n \text{is in bin } k) \leq \binom{n}{t} p^t = \binom{n}{t} (1/n)^t \tag{9.7}$$

Note that in this set-up, we guarantee the balls of S_n fall into k, but there might be other balls falling into it; we do not use $(1-p)^{n-k}$ and thus this effectively become 1^{n-k} or the other balls are free agents: they can fall into k or not.

The difference between Eq.(9.6) and Eq.(9.7) is a 1.5. We proceed discarding 1.5 as follows.

Using inequality 4.3 we obtain from Eq.(9.7) the following.

$$P(Bin\ k \geq t) \leq \left(\frac{ne}{tn}\right)^t \leq \left(\frac{e}{t}\right)^t.$$

Assum $n \geq e^2$ implying a $\ln n \geq 2$. Furthermore choose $t = c\ln(n)$, with $c = e^2$. Then if $t = e^2 \ln(n)$ we have the

following. Note that $2c = 2e^2 \approx 14.778 > 14$

$$\begin{aligned} P(Bin\ k \geq t) &\leq \left(\frac{e}{t}\right)^{c\ln n} \\ &\leq \left(\frac{e}{e^2 \ln n}\right)^{c\ln n} \\ &\leq \left(\frac{1}{e\ln n}\right)^{c\ln n} \\ &\leq \left(\frac{1}{e}\right)^{c\ln n} \cdot \left(\frac{1}{\ln n}\right)^{c\ln n} \\ &\leq \frac{1}{n^c} \cdot \frac{1}{n^c} \\ &\leq \frac{1}{n^{2c}} \leq \frac{1}{n^{14}}. \end{aligned}$$

Then

$$\begin{aligned} P(\text{number of balls any bin } \geq t) &= P((\text{number of balls of bin } 1 \geq t) \\ &\quad \cup (\text{number of balls of bin } 2 \geq t) \\ &\quad \cup \quad \ldots \\ &\quad \cup (\text{number of balls of bin k} \geq t) \\ &\quad \cup \quad \ldots \\ &\quad \cup (\text{number of balls of bin n} \geq t)) \\ &\leq \sum_i P((\text{number of balls of bin k} \geq t) \\ &= n \cdot P((\text{number of balls of bin k} \geq t) \\ &= n \cdot P((\text{Bin k} \geq t) \\ &\leq n \cdot \frac{1}{n^{14}} \\ &= \cdot \frac{1}{n^{13}}, \end{aligned}$$

since $c = e^2$. This means $t = e^2 \ln(n)$ is too generous. We can make it smaller.

The bound $t = c\ln n/\ln\ln n$ for $c = 2e^2$ gives. First let's bring our bound in an exponential form that we did not do so before.

$$\begin{aligned} P(\text{number of balls any bin } \geq t) &\leq n \cdot P((\text{number of balls of bin k} \geq t) \\ &\leq n \cdot \frac{e^t}{t^t} \\ &\leq \exp(\ln n + t - t\ln t) \\ &\leq \frac{1}{n^6}. \end{aligned}$$

The bound holds for $n > 4000000 > e^{e^{e^1}}$. This is to make sure $\ln\ln\ln n/\ln\ln n$ is small enough. Thus the result has been proven w.h.p i.e. that with probability at least $1 - 1/n^6$ every bit has no more than $2e^2 \ln n/\ln\ln n = \Theta(\ln n/\ln\ln n)$ balls. ■

Exercise 765.

If we throw n balls into n bins, what is,
(a) the expected number of empty bins?
(b) the expected number of bins with exactly one ball?
(c) the expected number of bins with exactly two balls?

Solution.
Part (a) has been answered earlier for the general case. Set $N = n$ to obtain the answer in (a) below. For all the parts below, let X be a random variable that takes values 1 or 0 depending on whether a bin is empty (for part (a)), contains exactly one ball (for part (b)), or exactly two balls (for part (c)). Then let Y be a random variable that counts the number of empty bins for part (a), or the balls with exactly one/two ball(s) for parts (b) and (c) respectively. It is clear that $E(Y) = nE(X)$.
(a) Let's consider one bin and call it for convenience only, bin 1. The probability that a ball falls in it is $1/n$ (and therefore the probability that a ball does not fall in it is $1-1/n$). The probability that none of the n balls falls into this bin is thus $(1-1/n)^n$. Then for the X variable we get that $E(X) = (1-1/n)^n$. Summing for the n bins we get that Y is such that: $E(Y) = n(1-1/n)^n \leq n/e$.
(b) Again, for bin 1, we find the probability that exactly one ball falls into that bin. Among n balls, the probability that exactly one falls into bin 1 is equal to $\binom{n}{1}(1/n)(1-1/n)^{n-1} = (1-1/n)^{n-1}$. As in part (a), we get $E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) = (1-1/n)^{n-1}$, and therefore for Y (that gives the number of bins with exactly one ball) we find that $E(Y) = n(1-1/n)^{n-1} = \frac{n^2}{n-1}(1-1/n)^n \leq \frac{n^2}{e(n-1)}$.
(c) As in parts (a), (b), the probability that bin 1 has exactly two balls is now: $\binom{n}{2}(1/n)^2(1-1/n)^{n-2}$, and therefore the X (the random variable that take values 1 or 0 depending on whether a given bin has exactly two ball or not) has expectation $E(X) = \binom{n}{2}(1/n)^2(1-1/n)^{n-2}$. Then for Y, we get that $E(Y) = nE(X) = n\frac{n(n-1)}{2}\frac{1}{n^2}\frac{(1-1/n)^n}{(1-1/n)^2} \leq \frac{n^2}{2e(n-1)}$.

■

Exercise 766.

Let us throw t balls into n bins. What is the probability that bin 5 is empty?
Implicit in the statement above is that we have enough bins to allow for bin 5 (i.e. if we label them $1,2,,3,\ldots$ we have 5 bins or more)!

Solution.
There is nothing specific about bin 5. We just want to stress that the question is for a specific bin. The possible outcomes is n^t. We can write down an t-digit sequence, and every 'digit' is a value from 1 to n indicating a bin number (that identifies a signle bin). The number of outcomes is n^t as each one of the t digits take n values. The event we are interested in is the one where the t-digit sequence contains no 5 at all. We have $(n-1)^t$ such sequences (all possible digits other than 5). Thus the probability of bin 5 being empty is $(n-1)^t$ divided by n^t i.e.

$$\frac{(n-1)^t}{n^t} = \left(1-\frac{1}{n}\right)^t.$$

■

Exercise 767.
Let us throw t balls into n bins. What is the probability that bins 5, 8 is empty?
Implicit in the statement above is that we have enough bins to allow for bin 5, 8 (i.e. if we label them $1,2,,3,\ldots,5,\ldots,8,\ldots$), we have 8 bins or more!

Solution.
It is not difficult to conclude that

$$\frac{(n-2)^t}{n^t} = \left(1-\frac{2}{n}\right)^t.$$

■

Exercise 768.

Let us throw balls into n bins one by one. Let Y_i be a random variable that counts the number of balls needed to fill one more bin, the i-th bin (not necessarily bin i). Obviously $Y_1 = 1$, since before the throw of the first ball we assume all bins are empty.
Examine the properties of random variable Y, where $Y = \sum_{i=1}^{n} Y_i$.
The problem is also known as the coupon's collector problem. Consider also an alternative view let us call it baseball-player cards. Bob has decided to to build a collection baseball cards of n players. For that he purchases individual enveloped baseball cards; one card is contained in an opaque envelope. The chance that he gets one of any of the n players of his election is $1/n$. How many cards should he gather to make sure that he has at least one of its one of the n players?

Solution.
For $Y_1 = 1$. Random variable Y_2 follows a geometric distribution with $Y_2 \sim g(p_2)$, where $p_2 = (n-1)/n$. We can even say that $Y_1 \sim g(p_1)$, whre $p_1 = n/n = 1$. In general $p_i = (n-(i-1))/n$. The expected number of balls thrown to fill a second bin is $E[Y_2]$, etc to fill the i-th bin is $E[Y_i]$ and so on. Since Y_i $g(p_i)$ with $p_=(n-(i-1))/n$ the $E[Y_i] = 1/p_i$. Therefore, using also the linearity of expectation,

$$E(Y) = E(\sum_i Y_i) = \sum_i E(Y_i) = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n(1+1/2+\ldots+1/n)nH_n,$$

where H_n is the harmonic series of order n. $H_n = \ln(n) + \gamma$, where γ is Euler's constant. Furthermore,

$$\begin{aligned} Var(Y) = var(\sum_i Y_i) &= \sum_i Var(Y_i) \\ &= \sum_{i=1}^{n} (1-p_i)/p_i^2 \\ &\leq \sum_{i=1}^{n} 1/p_i^2 \\ &= \sum_{i=1}^{n} \frac{n^2}{(n-i+1)^2} \\ &= n^2 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2} \\ &= n^2 \sum_{i=1}^{n} \frac{1}{i^2} \\ &= n^2 \pi^2/6, \end{aligned}$$

utilizing the Basel problem's sum due to Euler that provides the following result.

$$\sum_{i=1} n \frac{1}{i^2} = \pi^2/6.$$

■

Exercise 769.

We have m balls and n bins.
(a) What is the probability that two balls fall into the same bin?
(b) What is the probability that bin 5 is empty? (Implicit in the statement above is that we have enough bins to allow for bin 5 to exist!)
(c) What is the expected number of (ball) collisions?
(d) What is the expected number of empty bins.
(e) What is the probability bin 5 has k balls? Show that it is $\leq 1/k!$ for $m = n$.
(f) What is the probability that there is an empty bin (whether it is 5 or some other bin or more than one)?
(g) What is the probability bin 5 has AT LEAST k balls?
(h) What is the probability that the maximum number of balls in a bin is at least k ?

Solution.
(a) Obviously it is $1/n$. Pick two arbitrary balls i and j. The throws are independent of each other. Let e_{ij} be the event that they are in the same bin.

$$P(e_{ij}) = \sum_k P(\text{Ball i into bin k, AND Ball j into bin k}) = \sum_k \frac{1}{n} \cdot \frac{1}{n} = \frac{n}{n^2} = 1/n.$$

(b) There is nothing specific about bin 5. We just want to stress that the question is for a specific bin. The possible outcomes is n^m. We can write down an m-digit sequence, and every 'digit' is a value from 1 to n indicating a bin number (that identifies a signle bin). The number of outcomes is n^m as each one of the m digits take n values. The event we are interested in is the one where the m-digit sequence contains no 5 at all. We have $(n-1)^m$ such sequences (all possible digits other than 5). Thus the probability of bin 5 being empty is $(n-1)^m$ divided by n^m i.e.

$$\frac{(n-1)^m}{n^m} = \left(1 - \frac{1}{n}\right)^m.$$

Note that $1 - 1/n \leq e^{-1/n}$ and therefore,

$$\left(1 - \frac{1}{n}\right)^m \leq e^{-m/n}.$$

If $n = m$ this is at most $1/e$. One also could prove the same using indicator random variables. Here let us define Y_5 being 1 if bin 5 is empty and 0 otherwise. Obviously

$$E(Y_5) = P(\text{bin 5 is empty}) = \prod_{j=1}^{m} (1 - P(\text{ball j fall into 5}) = \prod_{j=1}^{m} (1 - 1/n) = (1 - 1/n)^m$$

(c) This derives from (a). The probability of a collision $P(e_{ij})$ is $1/n$. Let X_{ij} be an indicator random variable that is 1 if there is a collision between i and j and 0 otherwise. Then

$$E(X_{ij}) = 1 \cdot P(e_{ij}) + 0 \cdot \text{Irrelevant} = P(e_{ij}).$$

The number of collisions X is given by

$$X = \sum_i \sum_{j, i \neq j} X_{ij}$$

Therefore,

$$E(X) = E(\sum_i \sum_{j, i \neq j} X_{ij}) = \sum_i \sum_{j, i \neq j} E(X_{ij}) = \frac{m(m-1)}{2} E(X_{ij}) = \frac{m(m-1)}{2} \frac{1}{n} = \frac{m(m-1)}{2n}.$$

(d) Directly from (b). The probability a bin is empty is $\left(1 - \frac{1}{n}\right)^m$. Let Y_j be an indicator variable for bin Y_j, and

let $Y = \sum_{j=1}^{n} Y_j$ counts the number of empty bins. Using the same trick-methodology from the previous step we have, after noting $E(Y_i) = E(Y_5)$ for all j.

$$E(Y) = \sum_{i=1}^{n} E(Y_i) = n \cdot \left(1 - \frac{1}{n}\right)^m$$

For the same reasons as in (b) this is at most $ne^{-m/n}$. And for $m = n$ it is at most n/e.
(e) This is a binomial distribution $B(m,p;k) = B(m,1/n;k)$.

$$B(m,1/n;k) = \binom{m}{k}\left(\frac{1}{n}\right)^k (1-1/n)^{m-k} \leq \binom{m}{k}\frac{1}{n^k} \leq \left(\frac{m}{n}\right)^k \frac{1}{k!}.$$

Obviously the last expression is at most $1/k!$ for $m = n$.
(f)

$$\begin{aligned} P(\text{some bin is empty} &= P(\text{bin 1 is empty} \cup \text{bin 2 is empty} \cup \ldots \cup \text{bin n is empty}) \\ &\leq \sum_i P(\text{bin i is empty}) \\ &\leq \sum_i (1-1/n)^m \\ &= n(1-1/n)^m. \end{aligned}$$

(g) Pick k out of the m balls heading for bin 5. There are $\binom{n}{k}$ choices. And the chosen k balls ARE heading to bin 5 with probability $1/n^k$. We do not care about the remaining balls (they might be heading to bin 5 or not). Therefore the probability bin 5 has k or more balls is overestimated to be

$$P(\text{bin 5 has at least k balls}) \leq \binom{m}{k} \cdot (1/n^k)$$

Not bad! Compare it to the bound of (e) prior to the last step there of upper bounding.
(h) Let X_i be the number of balls in bin i.

$$P(\max_i X_i \geq k) = P(\exists i : X_i \geq k) \leq \sum_i P(X_i \geq k) = nP(X_i \geq k).$$

■

Exercise 770.

We throw n identical balls into n distinct bins with equal probability. Let X_i be the number of balls in bin i. Obviously $0 \leq X_i \leq n$ and $n = \sum_i X_i$. Moreover, $X_i \sim B(n,p;k)$ with $p = 1/n$, $q = 1-p = 1-1/n$. Furthermore, $E(X_i) = np = n(1/n) = 1$ and $var(X_i) = npq = n(1/n)(1-1/n) = 1-1/n = 1-q$.
(a) What is the probability that a given bin i gets more than $2\sqrt{n}$ balls? Use Chebyshev and Chernoff or Hoeffding bounds.
(b) What is the probability that any bin gets more than $2\sqrt{n}$ balls or equivalenty every bin has no more than $2\sqrt{n}$ balls?

Solution.
(a.1) Let X_i be the r.v. associated with the number of balls in bin i. Noting that $E(X_i) = 1$, and using Markov's inequality we have the following.

$$\begin{aligned} P(X_i > 2\sqrt{n}) &= P(X_i \geq 1 + 2\sqrt{n}) \\ &= P(X_i \geq E(X_i) + 2\sqrt{n}) \\ &= P(X_i - E(X_i) \geq 2\sqrt{n}) \\ &= P((X_i - E(X_i))^2 \geq 4n) \\ &\leq E((X_i - E(X_i))^2)/4n \\ &\leq Var(X_i)/4n \\ &\leq (1-1/n)/4n \\ &\leq 1/4n. \end{aligned}$$

(a.2)

Lemma 9.3

Let X_i be independent random variables following the same distribution as random variable X, $X \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$. Then, for any r such that $p < r < 1$ we have the following, after we substiture $r = (1+\delta)p$, $\delta > 0$.

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} \tag{9.8}$$

The following Corollary can be obtained from Lemma 9.3. It provides a more tangible upper bound than the generic one of the Lemma.

Corollary 9.14

Let X_i be independent random variables following the same distribution as random variable X, $X \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$. Then, for any $\delta > 2e-1$. we have the following,

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq 2^{-(1+\delta)pn}. \tag{9.9}$$

We first utilize the Chernoff bound of Corollary 9.14. We pick $\delta = 2\sqrt{n} > 2e-1$ for sufficiently large n, and remind ourselves $pn = 1$.

$$P(X_i > 2\sqrt{n}) = P(X_i \geq 1 + 2\sqrt{n}) = P(X_i \geq np + 2\sqrt{n}) = P(X_i \geq (1+\delta)np) = 2^{-\delta pn} = 2^{-2\sqrt{n}}$$

We next pick Lemma 9.3 with $r = 2/\sqrt{n}$. The bound is an even better $\approx 1/n^{\sqrt{n}}$. Hoeffding provides the same bound through Theorem 9.42. Corollary 9.15 or Corollary 9.16 of Hoeffding's theorem provide a lesser bound of $\exp(-8n)$.

Theorem 9.42

Let X_i be an independent random variable, $i = 1, \ldots, n$. Let $S_n = \sum_i X_i$, $\bar{X} = S_n/n$ and $p = E(\bar{X})$. Then for any h such that $0 < h < 1 - p$ we have the following.

$$\begin{aligned} P(\bar{X} - p \geq h)) &= P(S_n \geq (p+h)n) \\ &\leq \exp(-D((p+h)||p)n) \\ &\leq \left[\left(\frac{p}{p+h} \right)^{p+h} \left(\frac{1-p}{1-p-h} \right)^{1-p-h} \right]^n . \end{aligned} \tag{9.10}$$

Corollary 9.15

Let X_i be an independent random variable, $i = 1, \ldots, n$, such that $a \leq X_i \leq b$. Let $S_n = \sum_i X_i$ and $E(S_n) = np$. Then for any δ such that $0 < \delta < (1-p)/p$ we have the following.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-2n^2\delta^2 p^2}{n(b-a)^2} \right). \tag{9.11}$$

Corollary 9.16 appears in [26] as 5.3 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 9.16

Let X_i be an independent random variable, $i = 1, \ldots, n$, such that $a \leq X_i \leq b$. Let $S_n = \sum_i X_i$ and $E(S_n) = np$. Then for any h such that $0 < h < n - np$ we have the following.

$$P(S_n \geq pn + h) \leq \exp\left(\frac{-2h^2}{\sum_{i=1}^n (b_i - a_i)^2} \right). \tag{9.12}$$

(b) Let E_i be the event that a given bin i gets more than $2\sqrt{n}$ balls? By part (a) we proved $P(E_i) \leq 1/(4n$. We want to calculate the probabibility of event E where E is the event that any bin get more than $2\sqrt{n}$ balls. Then

$$P(E) = P(E_1 \cup E_2 \cup \ldots \cup E_n) \leq \sum_i P(E_i) \leq nP(E_i)$$

Using Chebyshev's bound $nP(E_i) \leq n(1/4n) \leq 1/4$

One could prove the latter as follows. Let Y_i be a random variable that is 1 if bin i has more than $2\sqrt{n}$ balls and 0 otherwise. Then $E(Y_i) = 1 \cdot P(X_i > 2\sqrt{n}) \leq 1/(4n)$. Let us then define random variable Y such that $Y = \sum_i Y_i$ i.e. it counts the number of bins with more than $2\sqrt{n}$ balls. Obviously $E(Y) = nE(Y_i)$ $P(E)$ is $P(Y \geq 1)$. The latter can be computed through Markov's inequality

$$P(Y \geq 1) \leq E(Y) \leq nE(Y_i) \leq n(1/(4n)) = 1/4.$$

■

Exercise 771.

We pick random points P_i in the unit square i.e. $P_i = (x_i, y_i)$, with $-1/2 \leq x_i, y_i \leq +1/2$. We then check whether P_i is inside the circle C with center $(0,0)$ i.e. $x^2 + y^2 \leq 1$. Then $p_i = P(P_i \in C) = \pi/4$, as the probability that P_i falls into C is proportional to the area of the circle relative to that of the square. We define random variable X_i that is equal to one if $P_i \in C$, and 0 otherwise.
We generate n random points P_i. Let S_n be another random variable with $S_n = \sum_i X_i$. We have $X_i \sim b(p_i)$. It is $p_i = p = \pi/4$. Furthermore $E(S_n) = nE(X_i) = np = n\pi/4$. Consider then another random variable $Y_n = (4/n) \cdot S_n$. Then

$$E(Y_n) = (4/n)E(S_n) = (4/n)(n\pi/4) = \pi$$

Therefore $Y_n = (4/\pi)S_n \to \pi$.
Compute

$$P(|Y_n - E(Y_n)| \leq \varepsilon\pi)$$

for some $0 < \varepsilon < 1$. Use a Chebyshev inequality and a Chernoff bound.

Solution.
Let us find some properties of the relevant random variables.

$$E(Y_n) = E((4/n)S_n) = (4/n)E(S_n) = (4/n)(n\pi/4) = \pi.$$

The we find the variance of Y_n; we note $p = \pi/4$ and $q = 1 - p = 1 - \pi/4$.

$$Var(Y_n) = Var((4/n)S_n) = (16/n^2)Var(S_n) = (16/n^2)(npq) = 16pq/n.$$

We use Chebyshev's inequality below

$$\begin{aligned} P(|Y_n - E(Y_n)| \geq \varepsilon\pi) &\leq \frac{Var(Y_n)}{\varepsilon^2\pi^2} \\ &\leq \frac{16pq}{n\varepsilon^2\pi^2} \\ &\leq \frac{4-\pi}{n\pi\varepsilon^2}. \end{aligned}$$

Therfore,

$$P(|Y_n - E(Y_n)| \leq \varepsilon\pi) \geq 1 - P(|Y_n - E(Y_n)| \geq \varepsilon\pi) \geq 1 - \frac{4-\pi}{n\pi\varepsilon^2}.$$

We use Chernoff's bound below in the form of Corollary 9.17.

Corollary 9.17

Let X_i be independent random variables following the same distribution as random variable X, $X \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$. Then, for any δ such that $0 < \delta < 1$ we have the following,

$$P(|S_n - np| \geq \delta pn) \leq 2 \cdot \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{9.13}$$

Note that $pn = \pi n/4$ and we use $\delta = \varepsilon$.

$$\begin{aligned} P(|Y_n - E(Y_n)| \geq \varepsilon\pi) &= P(|\frac{4}{n}S_n - \frac{4}{n}E(S_n)| \geq \varepsilon\pi) \\ &= P(|S_n - E(S_n)| \geq n\varepsilon\pi/4) \\ &\leq 2\exp\left(-\frac{\varepsilon^2\pi n}{3 \cdot 4}\right) \\ &\leq 2\exp\left(-\frac{\delta^2\pi n}{12}\right). \end{aligned}$$

■

For $P(|Y_n - E(Y_n)| \leq \varepsilon\pi)$ we do as before.

9.30 Collecting coupons

Exercise 772.

Coupon Collector

Let $\{1,2,\ldots,n\}$ be a set of n cards (coupons). In a collection of m randomly drawn coupons how large should m be so that there is at least one instance of each one of the n coupons? We will show that $m=(1+\varepsilon)n\ln n$ for some $\varepsilon>0$.

Solution.

For any fixed $i\in\{1,2,\ldots,n\}$, the probability i is not chosen in m choices is given by p_i

$$p_i=\left(1-\frac{1}{n}\right)^m=\exp(-\frac{m}{n})$$

Then the probability $p_1\vee p_2\vee p_3\vee\ldots\vee p_n$ is at most $\sum_i p_i=n\exp(-\frac{m}{n})$. The latter probability for $m=(1+\varepsilon)n\ln n$.

$$n\exp(-\frac{m}{n})=n\exp(-(1+\varepsilon)\ln n)=1/n^{\varepsilon}.$$

Anothe way to view this problem is by introducing an indicator variable $X_i=1$ if coupon i is never drawn and $X_i=0$ otherwise. The number of coupons NOT drawn is $X=\sum_i X_i$. The expected number of coupons NOT drawn is

$$E[X]=E[\sum_i X_i]=\sum_i E[X_i]$$

But $E[X_i]=1\cdot P(X_i=1)=(1-1/n)^m$. Thus

$$E[X]=E[\sum_i X_i]=\sum_i E[X_i]=n(1-1/n)^m$$

For $m=(1+\varepsilon)n\ln n$ this becomes $E[X]=1/n^{\varepsilon}$. Therefore if $E[X]<1$ there is a way that all coupons have been drawn.

■

Exercise 773.

Coupons revisited. We have a box of coupons labeled 1 through n. We perform with replacement the following experiment. We draw a coupon. If it has not be drawn before and is labeled i we check label i on a sheet as drawn and put it back into the box. We keep on drawing until all label have been checked on the sheet. How many coupons m are we expected to draw before all n labels are checked on the sheet.

Solution.
Let X_i be a random variable that indicates the number of draws to check on the sheet the first coupon with label $label(i)$. Note that the coupon checked on the sheet is not necessarily i but say $label(i)$. Note that $X_1 = 1$ as the first coupon drawn from the box will have its label $label(1)$ marked on the sheet. The total number of coupons drawn would be $X = \sum_i X_i$, where $X_1 = 1$. Moreover after the $k-1$ coupon has been marked, we have drawn so far $X_1 + \ldots + X_{k-1}$ coupons (with replacement). In the box $k-1$ coupons have already be marked, but $(n-(k-1)$ have not $P(X_k = r)$ follows a geometric distribution where $r-1$ draws are to one of the $k-1$ already picked coupons with the $(k-1)$-th draw to a new (k-th coupon). Thus $X_k \sim G((n-k+1)/n)$.
Parenthesis: Let $Z \sum G(p)$, where G is the geometric distribution with parameter p. The expectation $E(Z)$ is given as follows.

$$E(Z) = \sum_i i(1-p)^{i-1} p = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Therefore, using $E(Z)$, we conclude $E(X_k) = \frac{n}{n-k+1}$, since $p = (n-(k-1))/n$ for $X_k \sum G(p)$. Then

$$E(X) = \sum_k E(X_k) = \sum_k \frac{n}{n-k+1} = n \sum_k \frac{1}{k} \approx n \ln n.$$

■

9.31 Random graphs

Exercise 774.

A random graph with edge probability p is a graph that is formed by flipping a coin with probability p and deciding to include an edge if it comes H and not include edge for a T outcome. Let us assume that we use a fair coin $p = q = 1 - p = 1/2$. Is the graph connected?

Solution.

If the graph is not connected there exist at least two subgraphs ("components") with no edges from one to the other. If one subgraph G_1 has i vertices and the other/others G_2 has $n-i$ vertices it means the $i(n-i)$ vertices between the two pieces are missing. The probability that this is the case for one possible partion of i and $n-i$ is $2^{-i(n-i)}$. This is the probability the graph is NOT connected for a given split. For each value of i from 1 to $n/2$ there are $\binom{n}{i}$ ways to pick the vertices of G_1 and the remaining $n-i$ vertices are of G_2. The probability the graph G is not connected is the probability of the union of those events which is at most the sum of those probabilities. Thus

$$p \leq \sum_{i=1}^{i=n/2} \binom{n}{i} 2^{-i(n-i)} \leq \sum_{i=1}^{i=n/2} n^i 2^{-i(n-i)} \leq \sum_{i=1}^{i=n/2} (n2^{-n+i})^i \leq \sum_{i=1}^{i=n/2} (n2^{-n/2})^i \leq n/2 \cdot n2^{-n/2} = n^2/2^{n/2}$$

For $n/2^{n/2} < 1$ i.e. $n > 5$, the sum is a geometric sequence. Being a bit sloppy at the end we realize that $p \to 0$ as $n \to \infty$. Thus the graph is almost always connected.

■

Exercise 775.

A random directed graph with edge probability p is a graph that is formed by flipping a coin with probability p and deciding to include an edge in one direction if it comes H and include the edge in the opposite direction for a T outcome. Let us assume that we use a biased coin with $p = a/(n-1)$ and thus $q = 1-p$. What is the probability that for vertex i there is some edge directed into node i?

Solution.
Let us call p_i this probability. There are $n-1$ other vertices (other than i). If all of them are directed OUT of i this occurs with probability $(1-a/(n-1))^{(n-1)} \approx e^{-a}$. Thus p_i is given by the following equation

$$p_i = 1 - e^{-a}$$

What is now the probability Q that this is true for EVERY vertex?

$$Q = \prod_i p_i \leq (1-e^{-a})^n$$

Obviously $Q \to 0$ as $n \to \infty$. node?

■

Exercise 776.
A tournament is a directed graph which has one edge between every pair of vertices in one or the other direction. Some tournament contains $n!/2^n$ Hamiltonian cycles.

Solution.
For a given permutation of the vertices the probability it is a Hamiltonian cycles is $1/2^n$. There are $n!$ permutations of n vertices, so the expected number of Hamiltonian cycles is $n!/2^n$.

■

Exercise 777.

What is the expected number of Hamiltonian cycles in the random graph with edge probability $p = a/(n-1)$?

Solution.
Let now i range over the $n!$ permutations of the n vertices of the graph. Let $X_i = 1$ if permutation i leads to a hamiltonian cycle, and $X_i = 0$ otherwise. Then

$$E[X_i] = (a/(n-1))^n$$

and

$$E[X] = E[\sum_i X_i] = n!(a/(n-1))^n \approx (n/e)^n (a/(n-1))^n \approx (a/e)^n 1/(1-1/n)^n \approx (a/e)^n \cdot e \geq (a/e)^n$$

■

Exercise 778.

Show that for sufficiently large n, a random $n \times n$ bipartite graph where each possible edge is present with probability .5 is very likely to have a perfect matching. (Hint: recall Hall's theorem to decide whether a bipartite graph has a perfect matching).

Solution.

Hall's theorem states that a bipartite graph has a perfect matching unless there is a subset of vertices A on the left such that $|A| > |R(A)|$. We shall show that the probability that any A has $|R(A)| \leq |A| - 1$ is small. For a particular A, $|R(A)| < |A|$ only if there are at least $n - |A| + 1$ vertices on the right to which there are no edges from A. The chance that this happens for some A is less than the sum of the chances it happens for any particular A. Thus

$$\begin{aligned}
\text{P(there is no p.m.)} &\leq \sum_{i=1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

Thus for sufficiently large n, the chance that a random $n \times n$ bipartite graph does not have a perfect macthing is very small.

■

Exercise 779.

Let $G(n,p),\ 0 \leq p \leq 1$, be an undirected graph on n labeled vertices, where each edge e (among the $n(n-1)/2$ possible edges on n vertices) is included in the graph with edge probability p, independently of any other edge. If

$$p = \frac{(1+\varepsilon)\ln n}{n},$$

where $\varepsilon > 0$, show that with high probability (i.e. with probability tending to 1 when $n \to \infty$), $G(n,p)$ has no vertex of degree less than 11.

Solution.
The probability that a given vertex has degree exactly k is equal to

$$\binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Summing for all values of k from 0 to 10 we get the probability that a vertex has degree less than 11. If we multiply by n we get an upper bound on the probability that some vertex has degree less than 11. It suffices to prove that the latter expression (call it P) is upper bounded by a function $\varepsilon(n)$ such that $\varepsilon(n) \to 0$, as $n \to \infty$. This can be proven as follows . We use $p = (1+\varepsilon)\log n/n$, $\varepsilon > 0$, and thus $(1-p)^n \leq e^{-pn} = \frac{1}{n^{1+\varepsilon}}$, while $1/(1-p) < 2$. We also use $\binom{n-1}{k} \leq n^k$.

$$\begin{aligned}
P &= n\sum_{k=0}^{10} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&\leq n(1-p)^n \sum_{k=0}^{10} n^k \frac{(1+\varepsilon)^k \log^k n}{n^k (1-p)^{k+1}} \\
&\leq n\frac{1}{n^{1+\varepsilon}} \sum_{k=0}^{10} (1+\varepsilon)^k \log^k n\ 2^{k+1} \\
&\leq \frac{11\ 2^{11}(1+\varepsilon)^{10}\log^{10} n}{n^{\varepsilon}} \\
&\to 0 \text{ when } n \to \infty.
\end{aligned}$$

where the last sum, was bounded above by 11 times its largest term. Note, that no matter how bad i overestimated, I got the desired result. ■

Exercise 780.
Let $G(n,p)$, $0 \leq p \leq 1$, be an undirected graph on n vertices where each edge e (among the possible $n(n-1)/2$ edges on n vertices) is included in the graph with edge probability p independent of any other edge. If $p = \frac{1}{2}$ show that:
(a) With high probability (i.e. with probability tending to 1 as $n \to \infty$) the maximum size of an independent set in $G(n, \frac{1}{2})$ is no more than $(4+\varepsilon)\log n$, for any $\varepsilon > 0$.
(b) Deduce then, that for some constant $c > 0$, with probability tending to 1 as $n \to \infty$,

$$\frac{cn}{\log n} \leq \gamma(G),$$

where $\gamma(G)$ is the chromatic number of G.

Solution.
The probability that an n vertexgraph with edge probability $\frac{1}{2}$ has a k independent set is at most

$$\binom{n}{k} 2^{-\binom{k}{2}} \leq n^k 2^{-\frac{k^2}{4}}$$

For $k = (4+\varepsilon)\lg n$, this means the chance there is an independent set larger than that is no more than

$$\begin{aligned} (n2^{-\frac{k}{4}})^k &\leq (nn^{-1-\frac{\varepsilon}{4}})^k \\ &= n^{-\varepsilon k} \end{aligned}$$

Which converges to 0 for all $\varepsilon > 0$.

For part (b), simply recall that the vertices of any color class in G must be an independent set. As no color is used more times than the size of the largest independent set, the fact that with high probability no independent set is larger than $5\lg n$ implies that with high probability

$$\gamma(G) \geq \frac{n}{5\lg n}.$$

■

Exercise 781.

Show that any bipartite graph $G(X \cup Y, E)$ ($|X| = |Y| = n$) with edge probability $p = \frac{1}{2}$ has a perfect matching with high probability (i.e. with probability tending to 1 as $n \to \infty$).

Solution.

Hall's theorem states that a bipartite graph has a perfect matching unless there is a subset of vertices A on the left such that $|A| > |R(A)|$. We shall show that the probability that any A has $|R(A)| \leq |A| - 1$ is small. For a particular A, $|R(A)| < |A|$ only if there are at least $n - |A| + 1$ vertices on the right to which there are no edges from A. The chance that this happens for some A is less than the sum of the chances it happens for any particular A. Thus

$$\begin{aligned}
\text{P(there is no p.m.)} &\leq \sum_{i=1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \sum_{i=n/2+1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

Thus for sufficiently large n, the chance that a random $n \times n$ bipartite graph does not have a perfect matching is very small.

■

Exercise 782.

For an undirected connected graph $G = (V,E)$, let $d(i,j)$ be the shortest path (for graphs with no weights, the length) between vertex i and vertex j . The diameter $(diam(G))$ of G is defined to be the maximum among all the shortest paths between any two vertices in G, i.e.

$$diam(G) = \max_{i,j \in V \, i \neq j} d(i,j)$$

Show that for an appropriate p all $G_{n,p}$ graphs have diameter 2 with high probability (i.e. with probability tending to 1 as $n \to \infty$).

Solution.

A graph has diameter 2 if between any two distinct points there exist either an edge or a path of length 2, therefore for the second case, for every two points x, z there exists a point y connected to both x, z. For $G_{n,p}$ (we'll fix p at the end of our discussion) graph y is connected to both x,z with probability p^2. We examine the complement of the problem, namely we'll try to find the probability (an upper bound) that there exists a set of two points not connected by a path of length 1 or 2. This is at most

$$\binom{n}{2}(1-p)(1-p^2)^{(n-2)} \leq \frac{n^2}{2}e^{-p^2(n-2)}$$

since, we can choose two points x,z in $\binom{n}{2}$ ways and no other point y (among the other $n-2$ points) is in the path between x,z with probability $(1-p)^{(n-2)}$. Therefore the previous expression is an upper bound on the probability that there exists two points not connected by a path of length 2 (note that we ignore the case that the graph has diameter 1 which occurs with probability $p^{O(n^2)}$ since this term is finally absorbed by the term shown above). Let $p = \sqrt{\frac{(2+\varepsilon)\log n}{n}}$. Then the prior expression is bounded above by

$$\frac{1}{n^{\varepsilon}} \to 0 \quad as \quad n \to \infty$$

and therefore the probability that the graph is of diameter 2 tends to 1 for large n.

■

Exercise 783.

One may view a tournament T_n on n vertices as a tournament on n players, where an edge exists between vertex i and vertex j in T_n, if player i beats (or outranks) player j. A tournament T_n has property S_k if for every k players $x_1,\ldots,x_k$ there is some other player y who beats all of the other ones. Show that for every k, there is a finite T_n with property S_k. Find the smallest possible value of n you can, in terms of k, so that a tournament on at least so many vertices, has property S_k.

Solution.

We use a probabilistic argument to show that for n sufficiently large, the probability a random tournament does not have property S_k is less than 1. This probability is at most

$$\binom{n}{k}\Pr(\text{a particular } x_1,\ldots,x_k \text{ are not all beaten by some vertex} y)$$

For any set of k vertices, the probability some other particular vertex "beats" them all is 2^{-k}. So the chance it does not beat them all is $1-2^{-k}$, and the chance that none of the other vertices beat the entire set of k is $(1-2^{-k})^{n-k}$. Thus the probability that a random tournament does not have property S_k is at most

$$\binom{n}{k}\left(1-\frac{1}{2^k}\right)^{n-k} \quad < \quad n^k e^{-\frac{n-k}{2^k}}$$

For this to be less than 1, we want $e^{\frac{n-k}{2^k}} \geq n^k$, or equivalently $\frac{n-k}{2^k} \geq k\ln n$, and hence $n \geq k2^k \ln n + k$. This can be achieved by picking $n \geq 2k^2 2^k$.

■

Exercise 784.

Let $G_{n,p}$ be a undirected random graph on n vertices generated by independently including each edge with probability p. What is the expected number of (exactly) ten vertexcycles if $p = \frac{2}{n}$? How does this expectation grow with n (as n goes to ∞)?

Solution.

We first find the number of cycles we can form with k fixed objects (we will fix k to be 10 at the end). There are $k!$ permutations among k objects, and $k!/k = (k-1)!$ cyclic permutations among them (snce any cyclic permutation gives rise to k ordinary permutations) and these cyclic permutations define $(k-1)!/2$ unique cycles (since a cyclic permutation and its reverse define the same cycle, e.g. $1 \to 2 \to 3 \to 1 \equiv 3 \to 2 \to 1 \to 3$). We can choose k among n vertices in $\binom{n}{k}$ ways. From a cyclic permutation we get a cycle in a graph if all the k edges implied by the cyclic permutation appear, and this occurs with probability p^k. For $k = 10$ we have $\binom{n}{10}(9!/2)$ distinct cycles of length 10 and each of them appears with probability p^{10}, therefore the expected number of cycles of length 10 is:

$$E = \binom{n}{10}\frac{9!}{2}p^k = \frac{n!}{10!(n-10)!}\frac{9!}{2}(\frac{2}{n})^{10} = \frac{n!}{n^{10}(n-10)!}\frac{2^10}{2\cdot 10} = \frac{n(n-1)(n-2)\dots(n-9)}{n^{10}}\frac{2^10}{2\cdot 10}$$

where we substituted $p = 2/n$. For $n \to \infty$,

$$\frac{n(n-1)\dots(n-9)}{n^{10}} \to 1,$$

so we have

$$E \to \frac{2^{10}}{2\cdot 10} = 102.4$$

■

Exercise 785.

Suppose an n by n bipartite graph is generated randomly by including each edge independently with probability p.

(i) What is the expected number of perfect matchings?

(ii) For what value of p is this approximately 1?

(iii) For the value of p from ii), show that the probability that the graph has a perfect matching is exponentially small. (Hint: Consider the probability that there is an isolated vertex).

Solution.
(i) Every perfect matching in a bipartite graph $G = (X \cup Y, E)$ with $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$, corresponds to a permutation of the elements of X onto the elements of Y. We have $n!$ permutations (and each mapping of an element of X to an element of Y corresponds to an edge of a perfect matching) and each one of them has probability p^n to appear, and therefore the expected number of perfect matchings is $n!p^n$.
(ii) $n!p^n = 1$ gives $p \approx e/n$ (if we use Stirling's aprroximation for the factorial).
(iii) A given vertex of X is isolated with probability $(1-p)^n = (1-e/n)^n \approx e^{-e}$. The probability that a vertex is not isolated is $(1-e^{-e})$. The probability that all vertices of X are not isolated is $(1-e^{-e})^n$ (since all these probabilities are independent of each other). The probability that there exists a perfect matching that saturates X is at most the probability that all vertices of X are not isolated, and therefore this probability is at most $(1-e^{-e})^n$ which is exponentially small($1-e^{-e} < 1$ and therefore the $(1-e^{-e})^n \to 0$ exponentially fast).

■

Exercise 786.

Show that there exists an $n \times n$ matrix of 0's and 1's where each row has seven 1's, but where every $\frac{n}{2} \times \frac{n}{2}$ submatrix (a matrix made up of the intesections of subsets of $\frac{n}{2}$ of the rows and $\frac{n}{2}$ of the columns, not necessarily consecutive) contains at least one 1.

Solution.

We calculate the probability that a $\frac{n}{2} \times \frac{n}{2}$ submatrix contains only 0 entries. We examine the problem for the general case where we allow k 1's in a row and therefore the probability we have a 1 in a row is equal to $\frac{k}{n}$. The probability that a $\frac{n}{2} \times \frac{n}{2}$ submatrix contains only 0 entries is equal to $(1-\frac{k}{n})^{\frac{n^2}{4}}$ since a zero element appears in a row with probability $(1-\frac{k}{n})$ and an $\frac{n}{2} \times \frac{n}{2}$ submatrix has $n^2/4$ elements.

The number of $\frac{n}{2} \times \frac{n}{2}$ submatrices is equal to $\left(\begin{array}{c} n \\ \frac{n}{2} \end{array} \right)^2$ since we can choose $\frac{n}{2}$ rows (out of a total of n) in $\left(\begin{array}{c} n \\ \frac{n}{2} \end{array} \right)$ ways (the same holds for columns too).

Therefore, the probability that a $\frac{n}{2} \times \frac{n}{2}$ submatrix contains only 0 elements as entries is bounded above by

$$\left(\begin{array}{c} n \\ \frac{n}{2} \end{array} \right)^2 (1-\frac{k}{n})^{\frac{n^2}{4}}$$

We know employ Stirling's approximation formula. We get an upperbound by ignoring the square root terms.

$$\left(\begin{array}{c} n \\ \frac{n}{2} \end{array} \right)^2 (1-\frac{k}{n})^{\frac{n^2}{4}} \approx \left(\frac{(\frac{n}{e})^n}{(\frac{n}{2e})^{n/2}(\frac{n}{2e})^{n/2}} \right)^2 (1-\frac{k}{n})^{\frac{n}{k}\frac{kn}{4}} \leq 4^n e^{-\frac{kn}{4}} \tag{9.14}$$

since we know that

$$(1-\frac{k}{n})^{\frac{n}{k}} \leq e^{-1}$$

From the equation above we get that

$$4^n e^{-\frac{kn}{4}} = (\frac{4}{e^{k/4}})^n$$

For $k \geq 6$ (and therefore for k=7) the base $\frac{4}{e^{k/4}}$ is less than 1 and therefore equation (2) goes asymptotically to 0 when n goes to infinity. This means that the probability that a submatrix contains only 0 elements is small and therefore the probability that all submatrices contain at least one 1 is sufficiently large i.e. such a matrix exists. ■

Exercise 787.

Show that for sufficiently large n, there exists an $n \times n$ matrix of 0's and 1's where each pair of rows differs in at least $49n/100$ positions.

Solution.
The number of possible $n \times n$ matrices is 2^{n^2}. We examine the complement of the problem i.e we would like to have, for sufficiently large n, that the number of matrices with a pair of rows (at least) that differ in less than $\frac{49n}{100}$ positions is too small (in our discussion below we would ignore constant ± 1 differences in positions of the two rows). We now estimate the number of matrices that have a pair of rows with at most $\frac{49n}{100}$ differing positions (if we wanted to be correct this should be $\frac{49n}{100} - 1$ differing positions, but as we said we ignore such constants by overestimating). An overestimate on the number of such matrices is the following

$$\binom{n}{2} \sum_{i=0}^{\frac{49n}{100}} \binom{n}{i} \cdot 2^i \cdot 2^{n-i} \cdot 2^{n^2-2n}.$$

The first term gives the choices of 2 rows out of n. The last term gives the number of ways of filling the other $n-2$ rows of the matrix. We now explain the terms in the sum. The first term gives the number of ways of selecting i differing positions, the second term gives the number of ways of filling these positions. Note, that if we fix these i positions in the first row, we have the same positions in the second row fixed too (a 0 in one of these i positions in the first row is a 1 in the corresponding position in the second row and similarly for a 1 in the first row in one of these positions). The third term counts the number of ways of filling the $n-i$ identical positions (since they must be the same in the two chosen rows). Now we divide this expression by 2^{n^2} to find the probability that we have such a situation. Note that the sum is a geometric series and therefore the dominating term is $\binom{n}{\frac{49n}{100}}$. Since $2^i \cdot 2^{n-i} = 2^n$ we can move 2^n out of the sum, which is at most twice the dominating term.

$$\frac{\binom{n}{2} \cdot 2 \cdot \binom{n}{\frac{49n}{100}} \cdot 2^n \cdot 2^{n^2-2n}}{2^{n^2}} = \frac{\binom{n}{2} \cdot \binom{n}{\frac{49n}{100}}}{2^{n-1}}$$

For the $\binom{n}{\frac{49n}{100}}$ we use Stirling's approximation formula which finally yields (after upperbounding the square root that appears in the denominator) the following expression

$$\frac{\binom{n}{2} \cdot 2^{\alpha n}}{2^{n-1}}$$

where

$$a = -\frac{49}{100} \lg \frac{49}{100} - \frac{51}{100} \lg \frac{51}{100}.$$

All logarithms are base 2. The term $2^{\alpha n}$ is the result of writing as powers of 2 the terms $(49/100)^{49/100n}$, $(51/100)^{51/100n}$ that appear in the approximation of the $\binom{n}{\frac{49n}{100}}$ term since $b^{bn} = 2^{b \lg b \cdot n}$. Finally, we get that the ratio of "bad" matrices over the total number of 0-1 matrices is equal to

$$\frac{n^2}{2^{(1-\alpha)n-1}}$$

The nominator and denominator in the above term become equal for $n \approx 116660$ and therefore for n sufficiently large (say $n \geq 150000$, where this ratio is equal to 0.002) we have that the probability of having a "bad" matrix goes to 0, therefore the result we want to prove holds with high probability. The result, generally holds when instead of $\frac{49}{100}$ we have $\frac{1}{2} - \varepsilon$ for ε sufficiently small and positive.

■

Exercise 788.

What is the expected value of the *determinant* and the *permanent* of a $n \times n$ matrix if each element takes values 0 or 1 independently with probability a half?
The permanent of a matrix $A = [a_{ij}]$ is defined to be equal to:

$$per(A) = \sum_{\text{all permutations } \sigma \text{ on } \{1,2,\ldots,n\}} \prod_{i=1}^{n} a_{i\sigma(i)}$$

Solution.
We examine the case where $n > 1$, since the $n = 1$ is trivial (expectation $1/2$). Every a_{ij} element takes equiprobably only two values. The following trivially holds. $\prod_{i=1}^{n} a_{i\sigma(i)} = 1$ if and only if $a_{i\sigma(i)} = 1 \quad \forall i$. The probability that all $a_{i\sigma(i)} = 1, \forall i$ is just 2^{-n}, and the result holds for all permutations σ. Then:

$$E(\prod_{i=1}^{n} a_{i\sigma(i)}) = 1 \cdot P(\prod_{i=1}^{n} a_{i\sigma(i)} = 1) + 0 \cdot P(\prod_{i=1}^{n} a_{i\sigma(i)} = 0) = 1 \cdot 2^{-n} = 2^{-n} \tag{9.15}$$

We first find the expected value of the permanent. We have that:

$$E(per(A)) = E(\sum_{\text{all permutations } \sigma \text{ on } \{1,2,\ldots,n\}} \prod_{i=1}^{n} a_{i\sigma(i)}) = \sum_{\text{all } \sigma \text{ on } \{1,2,\ldots,n\}} E(\prod_{i=1}^{n} a_{i\sigma(i)})$$

since the expectation of the sum is equal to the sum of the expectations. The variable $\prod_{i=1}^{n} a_{i\sigma(i)}$ is equal to 1 iff $\forall i a_{i\sigma(i)} = 1$, that is with probability $\frac{1}{2^n}$. Otherwise, it is 0, and therfore, $\frac{1}{2^n}$ is also the expectation of $\prod_{i=1}^{n} a_{i\sigma(i)}$ for each of the $n!$ distinct permutations σ). Thus we get,

$$E(per(A)) = \sum_{\sigma \text{ on } \{1,2,\ldots,n\}} 2^{-n} = n! \cdot 2^{-n}$$

The computation of the determinant is very similar, useing the same observation as above. We need only to note that half the permutations are odd (and thus of sign 1) and half even (and of sign 0). (The definition of an odd or an even permutation can be found in the textbook.) Then $(-1)^{sign(\sigma)}$ is half of the times 1 and the other half -1.

$$E(det(A)) = \sum_{\sigma \text{ on } \{1,2,\ldots,n\}} (-1)^{sign(\sigma)} E(\prod_{i=1}^{n} a_{i\sigma(i)})) = \sum_{\sigma \text{ on } \{1,2,\ldots,n\}} (-1)^{sign(\sigma)} 2^{-n} = 2^{-n}\ 0 = 0$$

We got the first equality by using the formula for the expectation of the sum of random variables, and the observation that the value of the product does not depend on the permutation chosen but only on the values of the a_{ij}. The second equality is due to (1), and the third comes from the fact that $n!/2$ permutations are even and contribute each one of them an 1, while $n!/2$ are odd and contribute -1. Another way to solve this problem is expansion by minors.

Note. In a random bipartite graph (we have two sets of n vertices, the left and the right one, and a vertex of the left set is connected to a vertex of the right set with probability p, independently of the other choices), we show that the expected number of perfect matchings is $p^n n!$. The number of perfect matchings is just the permanent of a matrix with entries a_{ij} that take values 1 or 0 with probabilitites p and $1-p$ respectively (and a value of 1 indicates an edge from vertex i of the left set to vertex j of the right one). A permutation is just an 1-1 function on $\{1,2,\ldots,n\}$. Each permutation gives a perfect matching, and a given permutation appears with probability p^n (since an edge from vertex i of the left set to $\sigma(i)$ of the right set, for a permutation σ, appears with probability p). We have $n!$ permutations, therefore the expected number of perfect matchings is just $n!p^n$. Take $p = 1/2$ and you have another solution for the *permanent* part of the problem. ■

9.32 Puzzles

Exercise 789.

Alice and Bob have each received a sealed envelope and an assurance that each contains some money and that one contains exactly twice as much as the other. They are given the option of making the following agreement: they both open their envelopes and whoever has the more money gives it to the other person.
Alice convinces herself that taking up this option is advantageous to her by the following argument: Assume her envelope contains x amount of money. Then Bob's envelope contains either $2x$ or $x/2$, each possibility being with probability one half. Hence Alice's expected gain in taking up the option is

$$\frac{1}{2} \cdot 2x - \frac{1}{2} \cdot x = x/2 > 0$$

Since she has a positive expected gain it is worth her while to play the game. By an analogous argument Bob also concludes that taking up the option gives him an expected gain. Surely this is a contradiction.
Identify the fallacy in the previous paragraph.

Solution.
Alice and Bob are mistaken in assuming that the probabilities that the other envelope contains $2x$ and $x/2$ are both one half. That depends on the distribution of how the envelopes were originally filled, which is unknown and by no means necessarily uniform. So for different values of x, the probability that x is the larger of the two amounts is likely to vary, in which case this expected value calculation is not correct. ■

9.33 Ramsey's numbers

Exercise 790.

Let $r = R(C_3, k)$ be the smallest number of vertices so that no matter how the edges of K_r are colored using k colors, K_r has a monochromatic (i.e. all edges are of the same color) C_3 (a cycle of length 3) as a subgraph. Show, by using constructive rather than probabilistic arguments, that:

$$2^k \leq R(C_3, k) \leq 3 \cdot k!$$

Solution.

We first prove the lower bound through the use of induction (over k)! We are going to prove that $R(C_3, k) > 2^k$. For $k = 1$ the result is trivially true ($R(C_3, 1) = 3$). Suppose it is true for all values up to k. We are going to prove the bound for $k+1$, namely that $R(C_3, k) \geq 2^{k+1}$. Since we assume the result true for k, we pick two copies of the complete graph on 2^k vertices. We can color the first copy with k colors without a monochromatic triangle from the inductive hypothesis, and similarly the second copy using the same k colors. Then we can color the edges that go from the first copy of K_{2^k} to the second one with a new, $k+1$-st color. Since no triangle exists in any of the two copies for the first k colors then we can't have a triangle of the $k+1$-st color (this would require an edge of $k+1$-st color in a single copy of K_{2^k}) and thus, we get that the complete graph on $K^{2^{k+1}}$ is free of monochromatic triangles. We now prove the upper bound.

■

Exercise 791.

Show that the bound of the previous problem can be as small as $ek!+1$.

Solution.

We prove that $R(C_3,k) \leq 3k!$ by induction. It is trivially true that $R(C_3,1) \leq 3$. Assume it is true for all colors less than or equal to $k-1$. We will prove the claim for k colors (and let us use colors $1,\ldots,k$). Let us pick a complete graph on $3k!$ vertices. Color its edges in some way. Pick an arbitrary vertex and let's call it v. Let S_i be the vertices w that are connected to v thru an edge of color i. We have k colors and $3k!-1$ vertices adjacent to v. Then there must exist a set S_j for some color j of size at least $3(k-1)!$ (because if all sets had sizes at most $3(k-1)!-1$, then we would have had at most $3k!-k$ vertices adjacent to v, but for $k>1$, we have $3k!-1$ of them). If in S_j we can find two vertices u, w connected by an edge of color j, then we are done, a monochromatic triangle is (v,u,w). Otherwise no edge of color j connects any two vertices in set S_j, i.e. the edges of this set utilize the other $k-1$ colors only and the size of set S_j is at least $3(k-1)!$. We apply the inductive hypothesis on this set and we can find a monochromatic triangle in S_j. This completes the induction. ■

Exercise 792.

For any $k > 4$, show $R(k,k) \geq 2^{k/2}$.

Solution.

Let $n = R(k,k) < 2^{k/2}$.

We (uniformly at random) color the edges of K_n with red or blue color. Edges are colored independently of each other $P(red) = P(blue) = p = 1/2 = q = 1 - p$. Think of flipping a coin for every edge and if it comes H we interpret it as red and we interpret T for blue. Or we interpret H for an edge to include and likewise a T for an edge not to include

For every fixed set of k vertices, the probability that they form a clique (are all red) is $p = 2^{-\binom{k}{2}}$.

Likewise for every fixed set of k vertices, the probability that they form an independent set (are all blue) is also $p = 2^{-\binom{k}{2}}$.

There are $\binom{n}{k}$ k-sets of vertices that can give rise to a clique or an independent set. If we use Lemma 9.1 the probability of a union of events is at most their sum of probabilities. Thus

$$P(\ \textbf{Graph has k-clique or k independent set}\ \leq 2\binom{n}{k}2^{-\binom{k}{2}}.$$

Noting $n < 2^{k/2}$ we have the following

$$\begin{aligned}
2\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}} &= (\frac{ne}{k})^k(\frac{1}{2})^{k(k-1)/2} \\
&= \left(\frac{2^{k/2}e}{k}\right)^k\left(\frac{1}{2}\right)^{k(k-1)/2} \\
&= \left(\frac{2^{k/2}e}{k2^{(k-1)/2}}\right)^k \\
&= \left(\frac{e}{k2^{-1/2}}\right)^k \\
&= \left(\frac{\sqrt{2}e}{k}\right)^k
\end{aligned}$$

For $k > 4$ the probability is less than one. Thus there are graphs that contain neither a k-clique or k-independent set. This implies $R(k,k) > 2^{k/2}$.

■

Exercise 793.

Show that there is a constant $c > 0$ such that for all k, $R(3,k) \geq \frac{ck}{\ln k}$. Find a similar lower bound for $R(4,k)$. (Hint: Use a random graph argument with edge probability p appropriately chosen.)

Solution.

The stated bound for $R(3,k)$ is trivial since $n \leq k$. But to find the bound for $R(4,k)$ we will use a probabilistic argument that we present first for the $R(3,k)$ case.

We shall prove a lower bound for $R(3,k)$, by showing that for lesser n, the chance a random graph with edge probability p has either a 3-clique or k-independent set is less than 1. Such a result (for any value of p) means there must exist some graph that has neither. This probability is bounded above by

$$\binom{n}{3}p^3 + \binom{n}{k}(1-p)^{\binom{k}{2}} \leq \frac{n^3}{6}p^3 + n^k(1-p)^{\frac{k^2}{4}}$$

If we pick $p = n^{-1}$, then the first term becomes $1/6$, and we need only prove that the second is less than $5/6$. To do this observe that each each of the following relations implies the one above it.

$$\begin{aligned}
n^k(1-\frac{1}{n})^{\frac{k^2}{4}} &\leq \frac{5}{6} \\
n^k(1-\frac{1}{n})^{\frac{k^2 n}{4n}} &\leq \frac{5}{6} \\
n^k e^{-\frac{k^2}{4n}} &\leq \frac{5}{6} \\
k\ln n + \ln\frac{6}{5} &\leq \frac{k^2}{4n} \\
\ln n &\leq \frac{k}{n} \\
\ln c + \ln k - \ln\ln k &\leq \frac{\ln k}{c}
\end{aligned}$$

which is certainly true for $c = \frac{1}{2}$.

For $R(4,k)$ we get the similar expression of

$$\binom{n}{4}p^{\binom{4}{2}} + \binom{n}{k}(1-p)^{\binom{k}{2}} \leq \frac{n^4}{24}p^6 + n^k(1-p)^{\frac{k^2}{4}}$$

bounding the probability, and instead chose $p = n^{-\frac{2}{3}}$. Following the above reasoning, it suffices to show that

$$\begin{aligned}
\ln n &\leq kn^{-\frac{2}{3}}, \text{ or} \\
k &\geq n^{\frac{2}{3}}\ln n
\end{aligned}$$

To do this we can let $n = (\frac{ck}{\ln k})^{\frac{3}{2}}$ (with $c = .5$ works), thus improving our bound to be non-trivial. ■

Exercise 794.

The Noisy Ramsey number $NR(k,k)$ is defined to be the minimal number of vertices a graph can have so that it has a k-noisy-clique or a k-noisy-independent set. A k-noisy-clique is defined to be what we get from a k-clique if we delete $\frac{1}{10}k^2$ edges. A k-noisy-independent set is a set of k vertices joined by $\frac{1}{10}k^2$ edges (or if we delete so many edges, when no multiple edges are allowed, then we get a k-independent set). Show that $NR(k,k) \geq 2^{\varepsilon k}$, for some properly chosen ε.

Solution.
In order to find a lower bound for $NR(k,k)$ we show that for sufficiently small n, the probability that a random n vertexgraph has either a k-noisy independent set or a k-noisy clique is less than one. This means some graph must have neither. For a random n vertexgraph G,

$$\begin{aligned}
\Pr(G \text{ has a } k\text{-noisy clique}) &\leq \binom{n}{k} \Pr(\text{A particular set of } k \text{ vertices is a noisy clique}) \\
&\leq \binom{n}{k} \binom{\binom{k}{2}}{k^2/10} 2^{-\left(\binom{k}{2} - k^2/10\right)} \\
&\leq \binom{n}{k} \binom{k^2/2}{k^2/10} 2^{-\left(\frac{k(k-1)}{2} - \frac{k^2}{10}\right)} \\
&\leq n^k \binom{k^2/2}{k^2/10} 2^{-.4k^2 + k/2}
\end{aligned}$$

To bound this, observe the following (the second step uses a crude form of Stirling's approximation that $n! \geq (n/e)^n$).

$$\begin{aligned}
\binom{k^2/2}{k^2/10} &\leq \frac{\left(k^2/2\right)^{k^2/10}}{(k^2/10)!} \\
&\leq \frac{\left(k^2/2\right)^{k^2/10}}{(k^2/10e)^{k^2/10}} \\
&= (5e)^{k^2/10}
\end{aligned}$$

Using this result and substituting $2^{\varepsilon k}$ for n we get the upper bound

$$\Pr(G \text{ has a } k\text{-noisy clique}) \leq 2^{k^2\varepsilon} (5e)^{k^2/10} 2^{-.4k^2 + k/2}$$

Since this bound also applies for a k-noisy independent set, it suffices to chose ε such that this quantity is less than $1/2$. I.e. we must have

$$2^{k^2(\varepsilon - .4)} 2^{(k^2/10)\lg 5e} 2^{k/2} < \frac{1}{2}$$

This is true as long as $\varepsilon \leq .4 - \frac{1}{k^2} - \frac{1}{2k} - \frac{\lg 5e}{10} \approx .0235 - \frac{1}{k^2} - \frac{1}{2k}$. As k gets large, chosing $\varepsilon = .02$ suffices. (One could get a better bound by a more careful approximation). ■

9.34 Van der Waerden

Exercise 795.

(Van der Waerden's conjecture) Show that if $n < \sqrt{k}2^{\frac{k}{2}}$ then for some coloring of the integers $\{1,2,\ldots,n\}$ with two colors, neither color contains an arithmetic progression of length k.

Solution.

For a random coloring the probability that any particular arithmetic progression is monochromatic is $2^{-(k-1)}$ (once the first element is assigned a color, the remaining $k-1$ elements must be given the same one). If N is the number of such arithmetic progressions, and $N < 2^{k-1}$, the claim must be true since that implies an upper bound on the probability that a random coloring makes some progression monochromatic is less than one, which means there is some chance a random coloring makes no such progression monochromatic, which means there must be some particular coloring that fails.

To calculate the number of progressions we count the number of possible starting positions for each possible size for the gap between elements (i). This gap must be less than $n/(k-1)$ as otherwise the first and last elements would be separated by n or more places.

$$\begin{aligned}
N &\leq \sum_{i=1}^{\frac{n}{k-1}} (n - i(k-1)) \\
&= \frac{n^2}{k-1} - (k-1)\sum_{i=1}^{\frac{n}{k-1}} i \\
&= \frac{n^2}{k-1} - (k-1)\frac{\frac{n}{k-1}\frac{n+k-1}{k-1}}{2} \\
&= \frac{n^2}{k-1} - n\frac{n+k-1}{2(k-1)} \\
&= \frac{n^2}{k-1} - \frac{n^2}{2(k-1)} - \frac{n}{2} \\
&= \frac{n^2}{2(k-1)} - \frac{n}{2} \\
&< \frac{k2^k}{2(k-1)} - \frac{n}{2} \\
&< 2^{k-1}
\end{aligned}$$

The bound given is based on using 2^{-k} for the probability a progression is monochromatic, and bounding the number of progressions by $(n/k)n$. ■

Exercise 796.

For m, r, n positive integers the set $\{1,\ldots,m\}$ has property $B(r,n)$ if for some collection of r subsets of size n of $\{1,\ldots,m\}$ any 2-coloring of $\{1,\ldots,m\}$ results in some member of the collection being monochromatic. Find a lower bound on r such that $\{1,\ldots,m\}$ has property $B(r,n)$.

Solution.

The chance for a random coloring that a particular set of size n is monochromatic is $2^{-(n-1)}$. If we have r such sets, the chance that there exists one of them that is monochromatic is no more than $r2^{-(n-1)}$. Thus if $r < 2^{n-1}$, *any* collection of r sets of size n has some chance of having no monochromatic elements in a random coloring, which means there is some particular coloring for which it has no monochromatic elements. So $\{1,\ldots,m\}$ cannot have property $B(r,n)$ unless $r \geq 2^{n-1}$. ■

9.35 Hashing

Exercise 797.
(a) In hashing with chaining (uniform hashing assumption) for n keys and N slots, what is the probability that two keys collide?
(b) What is the exepcted number of collision for key x_i, $1 \leq i \leq n$?

Solution.
(a) From the ball and bin setup, the answer is $1/N$.
(b) Using an indicator random variable, $(n-1)/N$. ■

Exercise 798.

Hashing with chaining. A total of n keys are hashed into an $N = n$ slot table T. Answer the following.
(a) What is the expected number of keys per slot?
(b) Show that for $n \to \infty$, $P(\max_i n_i \geq 2\ln n) = 0$.
(c) What is the expected number of empty slots?
(d) What is the expected number of slots with one key?

Solution.
(a) Let n_i be the keys per slot i. Then n_i follows a binomial distribution with parameters n and $p = 1/n$ i.e. $B(n,p;k)$. Let $q = 1-p$, i.e. $p+q = 1$. Let n_i $B(n,p;k)$. Then $E[n_i]$.

$$\begin{aligned}
B(n,p;k) &= \binom{n}{k} p^k q^{n-k} \Rightarrow \\
E[n_i] &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k} \\
E[n_i] &= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k q^{n-k} \\
E[n_i] &= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
E[n_i] &= \sum_{k=1}^{n} \cdot \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \\
E[n_i] &= np \sum_{k=1}^{n} \cdot \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
E[n_i] &= np \sum_{k=1}^{n} \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} \\
E[n_i] &= np \sum_{k=0}^{n-1} \cdot \binom{n-1}{k} p^k q^{n-1-k} \\
E[n_i] &= np(p+q)^{n-1} \\
E[n_i] &= np
\end{aligned}$$

Another way to prove the same is the n keys follow a Bernoulli distribution with $p = 1/n$. Thus let X_j be 1 if key k_j falls into slot i and 0 otherwise. Then $n_i = \sum_j X_j$ and therefore

$$E[n_i] = E[\sum_j X_j] = nE[X_j] = np.$$

Theorem 9.43 **Chernoff**

Suppose $X_1, \ldots X_n$ are independent random variables taking values in $\{0,1\}$. Let $X = sum_j X_j$ denote their sum and let $m = E[X]$ denote the sum's expected value. Then for any $\delta > 0$,

$$P(X \geq (1+\delta)m) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^m$$

(b) For $\delta = 2\ln n - 1$, $m = np = n(1/n) = 1$, we have

$$\begin{aligned} P(n_i \geq 2\ln n) &\leq \frac{e^{\delta}}{(1+\delta)^{1+\delta}} \\ &\leq \frac{e^{2\ln n - 1}}{(2\ln n)^{2\ln n}} \\ &\leq \frac{n^2}{e \cdot n^2 \cdot n^{2\ln\ln n}} \\ &\leq \frac{1}{e \cdot n^{2\ln\ln n}} \end{aligned}$$

Then

$$\begin{aligned} P(max_i(n_i) \geq 2\ln n) &\leq nP(n_i \geq 2\ln n) \\ &\leq \frac{n}{e \cdot n^{2\ln\ln n}} \\ &\leq \frac{1}{n^{2\ln\ln n - 1}} \end{aligned}$$

The latter upper bound goes to 0 as $n \to \infty$. Moreover if $n > e^3$, we have $2\ln\ln n - 1 \geq 1.0$, and thus the probability is at most $1/n$.
(c) Since $n_i \sim B(n,p;k)$, we have that $P(n_i = 0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$. Let indicator random varible B_i be 1 if $n_i = 0$, and 0 otherwise. Then $C = \sum_i B_i$ is the number of empty slots.

$$E[C] = E[\sum_i B_i] = \sum_i E[B_i] = \sum_i P(n_i = 0) = n(1-p)^n$$

If $p = 1/n$ then $E[C] = n(1-1/n)^n \approx n/e$.
(d) Since $n_i \sim B(n,p;k)$, we have $P(n_i = 1) = np(1-p)^{n-1}$. Defining a random variable just like before and call D, the expected number of slots with one key, we derive similarly,

$$E[D] = n^2 p(1-p)^{n-1}$$

If $p = 1/n$ this is $E[D] = n(1-1/n)^{n-1} \approx n^2/(e(n-1))$. ■

Exercise 799.

Hashing with chaining. A total of n keys are hashed into an $N = n^2$ slot table T. Answer the following for simple uniform hashing (assumption).
(a) What is the probability that slot i is empty after hashing n keys?
(b) What is the probability that slot i has two or more keys?

Solution.
(a) The probability that slot i is empty is the probability that all n keys missed it. In general this is $(1-1/N)^n$. Probability $1/N$ is for a key k_j to hit i or $1-1/N$ is the probability that k_j misses i. Given that we have n keys the formula derives. One can prove the first derivation below by induction on n. It also follows by expanding the power on the left side

$$(1-\frac{1}{N})^n \quad \geq \quad (1-\frac{n}{N})$$

If $N > n$ then $\geq$ becomes $>$.
(b) This is a binomial distribution's term: $b(n,k;p)$, where $k=1$, $p=1/N$ and thus

$$\binom{n}{1}p^k(1-p)^{n-k} \quad = \quad \frac{n}{N}(1-1/N)^{n-1}$$

Given that $N \geq n$, we have $N > n-1$. Thus from the observation and deriviation in (a) we have

$$\begin{aligned}\frac{n}{N}(1-1/N)^{n-1} &> \frac{n}{N}\left(1-\frac{n-1}{N}\right)\\ &= \frac{n}{N}-\frac{n(n-1)}{N^2}\end{aligned}$$

(c) The probability that a slot has two or more keys is one minus the probability it has zero minus the probability it has one key. From (a) and (b) above we have

$$\begin{aligned}P(n_i \geq 2) &\leq 1-(1-\frac{n}{N})-\frac{n}{N}+\frac{n(n-1)}{N^2}\\ P(n_i \geq 2) &\leq \frac{n(n-1)}{N^2}\end{aligned}$$

If we substitute $N = n^2$ we see that the probability a slot i has 2 or more keys is at most $1/n^2$. The probability that there exists a slot (0 or 1 or etc or i or etc $N-1$) that has two or more keys is at most $n \cdot 1/n^2 = 1/n$. Equivalently all slots have zero or one keys with probability at least $1-1/n$. If we increase N from $N = n^2$ to $N = n^k$, $k > 2$ we can further decrease the bound $1/n$ or increase $1-1/n$. ■

Exercise 800.

Hashing with chaining continues with the following problem which also serves as a generalization. Assume uniform hashing i.e. a key x is equally likely to get hashed into any of the n slots. A total of n keys get hashed into n slots of table T. Let n_i be the number of keys hashed into slot i. Answer the following.
(a) What is the probability p_k that k keys get hashed into slot i?
(b) Let $k > 1$. Show that

$$p_k/p_{k+1} > 1.$$

(c) Show that for $k > 1$,

$$P(n_i \geq k) = \sum_{j=k}^{n} P(n_i = j) \leq nP(n_i = k).$$

(d) Show that for $n \geq 20$, $A = e^2 + 1$, and $k = A\frac{\ln n}{\ln\ln n}$,

$$P(n_i = k) \leq \frac{1}{n^7}.$$

(e) Conclude that $P(n_i \geq k) \leq \frac{1}{n^6}$.
(f) Provide an bound bound for the expected size of n_i, i.e. show that

$$E[n_i] \leq k+1,$$

where k is as defined in (d) above.
(g) After all keys are inserted into the slots randomly (uniformly at random), let D be the maximum number of keys in a slot. Show that q_k, the probability D is k is

$$q_k \leq np_k.$$

(h) Find the expected value of $E[D]$. (Hint: use the rationale of (f) above.)

Solution.
(a) The size of slot i, n_i, follows a binomial distribution that is, $n_i \sim B(n,p;k)$, where $p = 1/n$. Therefore,

$$p_k = P(n_i = k) = B(n, 1/n, k) = \binom{n}{k}(1/n)^k(1-1/n)^{n-k}.$$

(b) Let $k > np = n(1/n) = 1$. Consider p_k/p_{k+1}.

$$\frac{p_k}{p_{k+1}} = \frac{P(n_i = k)}{P(n_i = k+1)} = \frac{\binom{n}{k}(1/n)^k(1-1/n)^{n-k}}{\binom{n}{k+1}(1/n)^{k+1}(1-1/n)^{n-k-1}} = \frac{(k+1)(n-1)}{n-k}.$$

We claim that for $k > 1$, $(k+1)(n-1)/(n-k) > 1$. This is true as long as $kn > 1$. The latter is true as $k > 1$ and by implication $n \geq k > 1$ as well.
(c) We have that $P(n_i \geq k) = \sum_{j=k}^{n} P(n_i = j) = \sum_{j=k}^{n} p_j$. Since $p_k/p_{k+1} \geq 1$ by part (b), the terms p_j , $j \geq k$ are bounded above by p_k. Therefore

$$P(n_i \geq k) = \sum_{j=k}^{n} p_j \leq (n-k+1)p_k \leq np_k.$$

Moreover $p_k = P(n_i = k)$.
(d) From (a) above we have

$$p_k = \binom{n}{k}(1/n)^k(1-1/n)^{n-k},$$

and $(1-1/n)^{n-k} \leq 1$. Since $\binom{n}{k} \leq (ne/k)^k$, we then obtain the following.

$$\begin{aligned} p_k &\leq \binom{n}{k}(1/n)^k \\ &\leq (ne/k)^k(1/n)^k \\ &\leq e^k/k^k \\ &\leq \exp\left(\ln(e^k/k^k)\right) \\ &\leq \exp(\ln Z), \end{aligned}$$

where $Z = \ln(e^k/k^k)$. In the remainder we are going to first form $\ln Z = \ln(e^k/k^k) = k - k\ln k$. Then we are going to find an upper bound for $\ln Z \leq -7\ln n$, which would then translate into an upper bound for $\exp(\ln Z)$. This way we shall prove the following.

$$\begin{aligned} p_k &\leq \exp(\ln Z) \\ &\leq \exp(-7\ln n) \\ &\leq \frac{1}{n^7} \end{aligned}$$

We start with $\ln Z$ after choosing, $A = e^2 + 1$, and

$$k = A \cdot \frac{\ln n}{\ln\ln n}.$$

Then

$$\begin{aligned} \ln Z &= \ln(e^k/k^k) = k - k\ln k \\ &= \frac{A\ln n}{\ln\ln n} - \frac{A\ln n}{\ln\ln n} \cdot \ln\frac{A\ln n}{\ln\ln n} \\ &= \frac{A\ln n}{\ln\ln n} - \frac{A\ln n}{\ln\ln n} \cdot (\ln A + \ln\ln n - \ln\ln\ln n) \\ &= A(1-\ln A)\frac{\ln n}{\ln\ln n} - A\ln n + A\ln n \cdot \frac{\ln\ln\ln n}{\ln\ln n} \\ &= A(1-\ln A)\frac{\ln n}{\ln\ln n} - (A-1)\ln n + A\ln n \cdot \left(-1 + \frac{\ln\ln\ln n}{\ln\ln n}\right) \\ &= Q_1 - (A-1)\ln n + Q_2. \end{aligned}$$

Quantity Q_1 is such that

$$Q_1 = A(1-\ln A)\frac{\ln n}{\ln\ln n}.$$

Since $A = e^2 + 1$ and thus $\ln A > 2$, we conclude $Q_1 < 0$. Furthermore, quantity Q_2 is such that

$$Q_2 = A\ln n \cdot \left(-1 + \frac{\ln\ln\ln n}{\ln\ln n}\right).$$

$Q_2 < 0$ i.e. $1 \geq \frac{\ln\ln\ln n}{\ln\ln n}$ for $n \geq 20$. Therefore we have the following noting that $A = e^2 + 1$ and thus $A - 1 = e^2 \geq 7.38$.

$$\begin{aligned} \ln Z &= Q_1 - (A-1)\ln n + Q_2 \\ &\leq 0 - (A-1)\ln n + 0 \\ &\leq 0 - 7.38\ln n + 0 \\ &\leq 0 - 7\ln n + 0. \end{aligned}$$

We then conlude the desired result.

$$\begin{aligned} p_k &\leq \exp(\ln Z) \\ &\leq \exp(-7\ln n) \\ &\Leftrightarrow \\ p_k = P(n_i = k) &\leq \frac{1}{n^7}. \end{aligned}$$

(e) From (c) above, using the upper bound for p_k derived from (d) above we have that

$$P(n_i \geq k) \leq np_k \leq n\frac{1}{n^7} \leq \frac{1}{n^6}$$

(f) We are going to estimate $E[n_i]$ for any $i = 1, \ldots, n$. We have, note also from (b) that $p_k/p_{k+1} > 1$. The value of k used is

$$\begin{aligned} E[n_i] &= \sum_{j=0}^{n} j \cdot P(n_i = j) \\ &= \sum_{j=1}^{n} j \cdot P(n_i = j) \\ &= \sum_{j=1}^{k-1} j \cdot P(n_i = j) + \sum_{j=k}^{n} j \cdot P(n_i = j) \\ &\leq \sum_{j=1}^{k-1} k \cdot P(n_i = j) + \sum_{j=k}^{n} n \cdot P(n_i = j) \\ &\leq k \cdot \sum_{j=1}^{k-1} P(n_i = j) + n \cdot \sum_{j=k}^{n} \cdot P(n_i = j) \\ &\leq k \cdot 1 + n \cdot \sum_{j=k}^{n} \cdot P(n_i = k) \\ &\leq k + n \cdot n \cdot P(n_i = k) \end{aligned}$$

Combining with the result from (e) we conclude that

$$\begin{aligned} E[n_i] &\leq k + n \cdot n \cdot P(n_i = k) \\ &\leq k + n^2/n^6 < k + 1. \end{aligned}$$

The value of $A = e^2 + 1$ and k is given as described in (d) above.

$$k = A \cdot \frac{\ln n}{\ln \ln n}.$$

(g) Below k is as defined earlier in (d) above. Let $D = \max_{i=1}^{n} n_i = \max\{n_1, n_2, \ldots, n_n\}$. We utilize also (b) that

states $P(n_i = k) \leq 1/n^7$ and this is true for all i, $i = 1, \ldots, n$.

$$\begin{aligned} P(D = k) &= P(\max_{i=1}^{n} n_i = k) \\ &= P(\exists\ i : n_i = k \wedge n_l \leq k \forall l \neq i) \\ &\leq P(\exists\ i : n_i = k) \\ &\leq P(n_1 = k \cup n_2 = k \cup \ldots \cup n_n = k) \\ &\leq nP(n_i = k) \\ &\leq n \cdot (1/n^7) = \frac{1}{n^6}. \end{aligned}$$

(h) For the expected value of $E[D]$ we work similarly as before for the $E[n_i]$. The conclusion is the same $E[D] < k+1$. ■

9.36 Miscellanea

Exercise 801.

You have a fair die. You want to generate an integer 1 through 5 throwing this die uniformly at random.
(a) Show how to do it; make sure it always comes with an answer i.e. it terminates.
(b) Give the expected number of throws you require.
Now you have an unfair coin what over a toss it comes H (heads) with probability p, where $0 < p < 1$, and T (tails) with probability $1 - p$.
(c) How can you generate a fair toss for H (heads) and T (tails) with this unfair coin?
(d) Give the expected number of tosses you require.

Solution.
(a) Throw the die until it comes 1-5 i.e. ignore 6. If it comes x, where $x \neq 6$ report x. Otherwise continue.
(b) The number of throws n follows a geometric distribution with parameter $p = 5/6$. The expected number of throws is $1/p = 6/5$. i
(c) Toss the coin twice. It HT with probability $p(1-p)$ and you report T, or TH with probability $(1-p)p$ and you report H. If it comes HH or TT you ignore the outcomes. T and H reporting are equiprobable $p(1-p)$.
(d) Experiment of (c) resembles a geometric distribution with $P = 2p(1-p)$. Expectation of eventual success is $1/P = 0.5/(p(1-p))$. Moreover the number of coin tosses is $2 \cdot (1/P) = 1/(p(1-p))$ as each experiment involves two of them. ■

Exercise 802.

We have the following function call `RandomBit()` that returns a zero or one with equal probability $(1/2)$. We use if the as follows.

```
  flip(int n) {
1.  m = 0;
2.  for(i=0 ; i < n ; i++){
3.    m = 2 * m + RandomBit();
4.  }
5.  return(m)
```

(a) What is the minimum and the maximum value returned by m?
(b) Does `flip(n)` generate a uniformly at random drawn integer?

Solution.
(a) Obviously min m is 0 and max m is 2^n-1. One can prove it inductively. Prior to $i=0$, we have $m=0$. At the conclusion of $i=0$, the only possible values for RandomBit() are 0 and 1, and thus of m 0 or 1 respectively. If by induction at the conclusion of the $i=k$ iteration the min and max value of m are 0 and 2^k-1 respectively, then at the conclusion of the $i=k+1$ iteration m is minimally 0 (one more RandomBit() that is 0), and the maximum value is

$$m = 2*m+1 = 2*(2^k-1)+1 = 2^{k+1}-1.$$

Thus at the conclusion of $i=n-1$, we have that the maximum value of m is 2^n-1.
(b) When m is equal to N at the conclusion of iteration $n-1$ then, m has generated the righmost bit of N. In other words, m generates in iteration i the i-leftmost bit of N. The range of values for N is 0 and $2^0+\ldots+2n-1=2^n-1$, and each such value is generated uniformly at random with probability $\underbrace{1/2\times 1/2\times\ldots\times 1/2}_{n \text{ times}}$.

■

Exercise 803.

In order to find the minimum (MIN) of n keys of $A[0..n-1]$, the following algorithm can be used.

```
  Min(A[0..n-1],n)
1. min = A[0];
2. for(i=1 ; i < n ; i++){
3.   if A[i] < min
4.     min=A[i];
5. }
6. return(min);
```

Assume all elements of A are distinct
(a) What is the probability that $A[i]$ is the MIN?
(b) What is the probability that line 4 is executed?
(c) What is the expected number of times line 4 is executed?

Solution.
(a) The probability that $A[i]$ is the MIN is $1/n$, as there are n numbers distributed over the n slots of A.
(b) The probability that line 4 is executed is the probability that $A[i]$ is the minimum of the i keys $A[0],\ldots A[i]$ i.e. $1/(i+1)$ using the same argument as in (a) above.
(c) Let X_i be a random variable that is 1 if line 4 is executed and 0 otherwise. Then $X=\sum_i X_i$ is the number of times line 4 is executed. We are interested in finding

$$E[X]=E[\sum_i X_i]=\sum_i E[X_i]$$

We note that

$$E[X_i]=1\cdot P(X_i=1)+0\cdot P(X_i=0)=P(X_i=1)$$

From part (b) P ($X_i=1$) $=1/(i+1)$. Therefore,

$$E[X]=\sum_i E[X_i]=\sum_{i=0}^{n-1}\frac{1}{i+1}.$$

The latter sum is the harmonic series H_n thus

$$E[X]=\ln n+\gamma.$$

■

Exercise 804.

SomeIT has two types of professors: `mean` and `nice`. A generous professor gives A's to all 100% of the students in his class. A mean professor gives A's to 80% of the students and $B+$ to the remaining students. Let's assume that all course/class grades are available in the form of arrays and all arrays have the same size n. How fast can you determine reliably (i.e. you are confident on your estimate 51% of the time or more) whether the grades of a given professor correspond to a mean or a nice one?
Note that your algorithm does not need to be correct all the time. However it needs to be fast. An obvious $O(n)$ solution that is correct 100% of the time is to go through array G and if we find a $B+$ we return `mean` otherwise we return `generous`.
This is too slow however.

Solution.
We look at say 11 random indices $i_1, \ldots i_{11}$.

```
1. SomeIT(A.roster,n,A.instructor)
2.  for(j=1;j<=11;j++){
3.    k=ChooseRandom(1,n) //Choose random index in 1..n
4.    i[j]=k;
5.    if A[k] == "B+'
6.        printf("mean");
7.        return;
8.  }
9.  printf("nice");
10. return;
```

SomeIT is an instance of a probabilistic algorithm. It always returns an answer, but sometimes the answer might be incorrect (unreliable). When SomeIT returns `mean` through lines 6-7, we know that the instructor is indeed `"mean"`, since a `B+` grade has been found in the roster. When SomeIT returns `nice` through lines 9-10 we might have of two cases.
Case 1. The instructor in question is `nice` and there was no chance of finding a `B+` whether 11 or n of the grades were to be examined. The answer given is thus correct.
Case 2. The instructor in question is `mean` and thus the answer given is misleading and unreliable. SomeIT failed to detect a `B+` in 11 trials, i.e. it detected an `A` in 11 trials.
What are the chances that the answer is unreliable i.e. Case 2 applies? The chance that we get an `A` answer to a lookup given a `mean` professor is 0.8 (80%). The chance that we get an `A` 11 times in a row in randomly picked probes is $0.8^11 < 0.086 < 0.10$. Therefore $> 1 - 0.10 = 90\%$ of the time a `mean` professor would be detected through lines 6-7, but only $\leq 8.6\%$ of the time a potentially unreliable answer will be given through lines 9-10.
A probabilistic algorithm similar to SomeIT always tells the truth when it says `mean`. Every time it says `nice` the result is reliable with confidence 90% or more. SomeIT is what we call a `Monte-Carlo` algorithm. ■

Part VI

Theory of numbers

Chapter 10

Introduction to Number theory

10.1 Short review on number theory

10.1.1 Numbers

Definition 10.1 **Residue classes modulo n**

Let n be a positive integer. The set of all residue classes modulo n or congruences mod n (or modulo n) is denoted in simple form by $\mathbb{Z}/n\mathbb{Z}$.

$$\mathbb{Z}/n\mathbb{Z} = \{0, 1, 2, \ldots, n-1\}.$$

This is supposed to mean the following.

$$\mathbb{Z}/n\mathbb{Z} = \{(0)_n, (1)_n, (2)_n, \ldots, (n-1)_n\}.$$

We read this set " $\mathbb{Z}$ modulo n " or " $\mathbb{Z}$ mod n ". Sometimes we write $\mathbb{Z}_n$ instead of $\mathbb{Z}/n\mathbb{Z}$. The latter notation explains set $\mathbb{Z}$ is split by n into n equivalence classes. Two integer x, y belong to the same equivalence class if $(x-y)$ or $(y-x)$ is a multiple of n. The equivalence class of x is $x \bmod n$.

Integer one is an integer, a rational and algebraic and thus a real number. Number $4/5$ is rational and algebraic and thus a real number. Number $\sqrt{2}$ is irrational, algebraic and thus a real number. Number π is irrational, transcendental and thus a real number.

Definition 10.2 **Equivalence relation**

A binary relation R on X is an equivalence relation on X if for every $x, y, z \in X$ the following properties are satisfied.

- (a) Reflexivity xRx for every $x \in X$,
- (b) symmetry xRy then yRx, and
- (c) transitivity xRy and yRz then xRz.

Definition 10.3 **Equivalence classes**

In an equivalence relation we can decompose X into disjoint subsets known as equivalence classes. Then for each $x \in X$ the set of elements equivalent to x define

$$[x/R] = \{y \in X : xRy\}.$$

Then

$$p : X \to X/R \text{ with } p(x) = [x].$$

For an equivalence relation R if aRb the $[a] = [b]$. (It is the same as $[a/R] = [b/R]$.)

10.1.2 Binary operators and operations

Definition 10.4 **Binary operator**

A binary operator defines a binary operation.

A binary operation is the operation implied by a binary operator.

Definition 10.5 **Binary operation**

For a set S and a binary operator $\oplus$, a binary operation is a rule that maps to each pair $(x, y) \in S \times S$ a unique $z \in S$.

Operations $+$, $*$ are binary operations for a set $S = \mathbb{N}$ or $S = \mathbb{R}$. In all definitions below we assume S is non-empty.

10.1.3 Groupoids and semigroups

Definition 10.6 **Magma/Groupoid**

Let S be a set, and $\oplus$ be a closed operator $S \times S \mapsto S$ defining a total function as follows.

$$\forall (x, y) \in S \times S : (x, y) \mapsto x \oplus y \in S.$$

The structure S equipped with $\oplus$ as structured is a magma. We then say $(S, \oplus)$ defines a magma, also known as a groupoid.

Operator $\oplus$, and when introduced operator $\otimes$, are closed operators $S \times S \mapsto S$ defining a total function as follows.

$$\forall (x, y) \in S \times S : (x, y) \mapsto x \oplus y \in S.$$

This is also known as a closure property.

Definition 10.7 **Semigroup**

Let S be a non-empty set, and $\oplus$ be an operator $\oplus : S \times S \mapsto S$. If operator $\oplus$ has the closure property and is associative, then the structure S equipped with $\oplus$ as structured is a semigroup.

1. Closure property: $\forall x \in S, \forall y \in S : x \oplus y \in S$.
2. Associativity : $(x \oplus y) \oplus z = x \oplus (y \oplus z) \quad \forall x \in S, \forall y \in S, \forall z \in S,$

We then say equivalently that $(S, \oplus)$ defines a semigroup.

A monoid is a semigroup with an identity element.

Definition 10.8 **Monoid**

An $(S, \oplus)$ that is a semigroup it becomes a monoid if there exists an element e in S such that

$$\forall x \in S : e \oplus x = x \oplus e = x.$$

Then $(S, \oplus)$ is called a monoid.

A monoid equipped with an inverse (every element has an inverse element) is a group.

10.1.4 Groups

Definition 10.9 **Group**

An $(S,\oplus)$ that is a monoid it becomes a group if every element of S has an inverse element.

$$\forall x \in S, \exists y \in S : x \oplus y = y \oplus x = e,$$

where e is the identity element of the underlying monoid. Then $(S,\oplus)$ is called a group.

Below we give a standalone definition of a group.

Definition 10.10 **Group**

An $(S,\oplus)$ with S non-empty is a group if it has the following properties.

1. Closure property: $\forall x \in S, \forall y \in S : x \oplus y \in S$.
2. Associativity : $(x \oplus y) \oplus z = x \oplus (y \oplus z) \quad \forall x \in S, \forall y \in S, \forall z \in S,$
3. Identity: there exists an $e \in S$ such that $\forall x \in S : e \oplus x = x \oplus e = x$.
4. Inverse: $\forall x \in S, \exists y \in S : x \oplus y = y \oplus x = e$.

Then $(S,\oplus)$ is called a group.

Example 10.1.1

For $S = \mathbb{Z}$ and $\oplus = +$ we define $e = 0$. The inverse y of x is then denoted by $-x$. $(\mathbb{Z},+)$ is a group. It is also an Abelian group, defined next.

Definition 10.11 **Commutative or Abelian group**

An $(S,\oplus)$ that is a group becomes a commutative group if

$$\forall x \in S, \forall y \in S : x \oplus y = y \oplus x.$$

Then $(S,\oplus)$ is called a commutative group, also known as an Abelian group.

$(\mathbb{R},+)$ is also a commutative (abelian) group.

Example 10.1.2

For $S = \mathbb{R}$ and $\oplus = +$ we define $e = 0$ The inverse y of x is then denoted as $-x$. $(\mathbb{Z},+)$ is a group.

Example 10.1.3

For an $\oplus = +$ operator we use 0 as its identity element that is, $e = 0$. We denote the inverse of a as $-a$. For an $\oplus = *$ operator we use 1 as its identity element that is, $e = 1$. We denote the inverse of an $a \neq 0$ as a^{-1} or $1/a$.

A group with operator $\oplus = +$ is known as an additive group. A group with operator $\oplus = *$ is known as a multiplicative group.

Definition 10.12 **Order of a group**

The order of a group $(S,\oplus)$ is the number of its elements, and it is denoted $|S|$ (cardinality symbol).

Definition 10.13 **Subgroup**

For a group $(S,\oplus)$ and a non-empty subset T of S that is $T \subseteq S$, we say T a sugroup of S if $(T,\oplus)$ is a group.

The order of a group is the cardinality of its underlying set S.

$(\mathbb{R},+)$ is also an abelian group. $(\mathbb{Z}_+,+)$ is not an abelian group. There is no identity element then. $(\mathbb{Z}_+,*)$ is not an abelian group. There is an identity element then which is 1, but there is no inverse for most elements (other than one)! $(\mathbb{Q}_+,*)$ is an abelian group. $(\mathbb{R}_+,*)$ is an abelian group. $(\mathbb{Q}^*,*)$ is an abelian group. $(\mathbb{R}^*,*)$ is an abelian group.

Example 10.1.4

For $S = \mathbb{R} - \{0\} = \mathbb{R}^x$ and $\oplus = \cdot$ we define $e = 1$ The inverse y of x is then denoted as $1/x$. $(\mathbb{R}^x,\cdot)$ is a group.

$(\mathbb{R}^x,\cdot)$ is also an abelian group.

Definition 10.14 **Cyclic group**

For a group $(S,\otimes)$ and $a \in S$, the elements a^i, where i is an integer, form a subgroup of S, called the subgroup generated by a. A group is cyclic if there is an element $a \in S$ that the subgroup generated by a is S itself. We can the write S as follows.

$$S = \{e, a, a^2, \ldots, a^{k-1}\},$$

where e is the identity of the group, and k is the smallest (positive) integer such that $a^k = e$.

If S is infinite,

$$S = \{\ldots a^{-2}, a^{-1}, e, a, a^2, \ldots\},$$

10.1.5 Rings and Integral domains

Definition 10.15 **Ring**

A ring is an algebraic structure consisting of a set S in which two binary closed operations are defined say $\oplus, \otimes$. Then, $(S,\oplus,\otimes)$ is a ring if (a) $(S,\oplus)$ is an abelian group, (b) $(S,\otimes)$ is a semigroup, and (c) $\otimes$ is distributive over $\oplus$, with e being the identity element of $(S,\oplus)$. A ring denoted by $(S,\oplus,\otimes)$ has the following properties.

1. Closure property for $\oplus$: $\forall x \in S, \forall y \in S : x \oplus y \in S$.
2. Associativity for $\oplus$: $\forall x \in S, \forall y \in S, \forall z \in S, (x \oplus y) \oplus z = x \oplus (y \oplus z)$
3. Commutativity for $\oplus$: $\forall x \in S, \forall y \in S : x \oplus y = y \oplus x$.
4. Identity for $\oplus$: there exists an $e = 0 \in S$ such that $\forall x \in S : 0 \oplus x = x \oplus 0 = x$.
5. Additive inverse for $\oplus$: $\forall x \in S, \exists y = -x \in S : x \oplus y = y \oplus x = e$.
6. Closure property for $\otimes$: $\forall x \in S, \forall y \in S : x \otimes y \in S$.
7. Associativity for $\otimes$: $\forall x \in S, \forall y \in S, \forall z \in S, (x \otimes y) \otimes z = x \otimes (y \otimes z)$.
8. Distribution of $\otimes$ with respect to $\oplus$:
   $\forall x \in S, \forall y \in S, \forall z \in S : x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z) \wedge (x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$.

Then $(S,\oplus,\otimes)$ is called a ring.

All of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are rings; in fact they are commutative rings for $\otimes$ being commutative as well. Non-commutative rings are $n \times n$ matrices whose elements are from the rings above.

Definition 10.16 **Commutative ring**

$(S,\oplus,\otimes)$ is called a commutative ring if $(S,\oplus,\otimes)$ is a ring and $\otimes$ is also commutative.

$$\forall x \in S, \forall y \in S : x \otimes y = y \otimes x.$$

Definition 10.17 **Ring with identity**

$(S,\oplus,\otimes)$ is called a ring with identity if $(S,\oplus,\otimes)$ is a ring and $\otimes$ supports an identity element let us call it 1.

$$\text{Identity for } \otimes : \exists e = 1 \in S : \forall x \in S : 1 \otimes x = x \otimes 1 = x.$$

Definition 10.18 **Integral domain**

Let $(S,\oplus,\times)$ be a non-zero commutative ring with $0 = e \neq 1$, where e or 0 is the identity of $\oplus$ and 1 is the identity of $\otimes$, that has no zero divisors which is equivalent to the following.

$$a \neq 0, b \neq 0 \Rightarrow c = ab \neq 0 \text{ or equivalently } c = ab = 0 \Rightarrow a = 0 \vee b = 0.$$

Then the structure $(S,\oplus,\otimes)$ is called an integral domain.

Definition 10.19 **Division ring**

A division ring $(S,\oplus,\otimes)$ is a ring with identity $1 \neq 0$ with an inverse element over $\otimes$, that is

$$\forall x \in S, x \neq 0, \exists a \in S : ax = xa = 1.$$

Definition 10.20 **Ideal**

For a ring $(S,\oplus,\otimes)$, a left ideal I of S is a subset of S that is a subgroup of the structure $(S,\oplus)$ of the ring that is closed under left multiplication over elements of S. Thefore $e \in I$, $e = 0$ being the identity of $(S,\oplus)$, and for every $a,b \in I$ and $x \in S$ the following are true.

$$a + b \in I, \quad -a \in I, \quad xa \in I.$$

A right ideal is defined similarly. A two-sided ideal is a left and right ideal; oftentimes it is called an ideal. If the ring is commutative, then the left, right, and two-sided ideals coincide and the term ideal identifies all three of them. Then the ideal I is an abelian subgroup.

10.1.6 Fields

A field can be defined as an extension of a ring as follows.

Definition 10.21 **Field from ring**

A field $(S,\oplus,\otimes)$ is a division ring that supports a commutative $\otimes$.

$$\forall x \in S, \forall y \in S : x \otimes y = y \otimes x.$$

A field can also be defined as an extension of two groups as follows.

Definition 10.22 **Field from groups**

For a field $(S,\oplus,\otimes)$, $(S,\oplus)$ is an abelian group, $(S,\otimes)$ is an abelian group but with an inverse defined only for $S-\{0\}$. Moreover $\otimes$ is distributive with respect to $\oplus$.

An equivalent standalone definition of a field follows.

Definition 10.23 **Field**

$(S,\oplus,\otimes)$ is a field with two binary operators $\oplus,\otimes$ if S has at least two elements, there exist $1\neq 0$ where 0 is the identity of $\oplus$, and 1 is the identity of $\otimes$ and it supports the following properties.

1. Closure property for $\oplus$: $\forall x\in S,\forall y\in S: x\oplus y\in S$.
2. Associativity for $\oplus$: $\forall x\in S,\forall y\in S,\forall z\in S,(x\oplus y)\oplus z=x\oplus(y\oplus z)$
3. Commutativity for $\oplus$: $\forall x\in S,\forall y\in S: x\oplus y=y\oplus x$.
4. (Additive) Identity for $\oplus$: there exists an $e=0\in S$ such that $\forall x\in S: 0\oplus x=x\oplus 0=x$.
5. Additive inverse for $\oplus$: $\forall x\in S,\exists y=-x\in S: x\oplus y=y\oplus x=e$.
6. Closure property for $\otimes$: $\forall x\in S,\forall y\in S: x\otimes y\in S$.
7. Associativity for $\otimes$: $\forall x\in S,\forall y\in S,\forall z\in S,(x\otimes y)\otimes z=x\otimes(y\otimes z)$.
8. Commutativity for $\otimes$: $\forall x\in S,\forall y\in S: x\otimes y=y\otimes x$.
9. (Multiplicative) Identity for $\otimes$: there exists an $e=1\in S$, $1\neq 0$ such that $\forall x\in S: x\otimes 1=1\otimes x=x$.
10. (Multiplicative) Inverse for $\otimes$: $\forall x\in S,\exists y=-x\in S: x\otimes y=y\otimes x=e$.
11. Distribution of $\otimes$ with respect to $\oplus$: $\forall x\in S,\forall y\in S,\forall zinS$:

$$x\otimes(y\oplus z)=(x\otimes y)\oplus(x\otimes z),$$

$$(x\oplus y)\otimes z=(x\otimes z)\oplus(y\otimes z).$$

Then $(S,\oplus,\otimes)$ is called a field.

Roughly speaking for a field $(S,\oplus,\otimes)$, $(S,\oplus)$ is an abelian group, $(S,\otimes)$ is short of an abelian group, with "short" defined that only elements $x\neq e$ i.e. $x\neq 0$ have a multiplicative inverse $y=1/x=x^{-1}$, that is, $(S-\{e\},\otimes)$ is a group.

Definition 10.24 **Finite field**

A finite field is a field that has a finite number of elements or equivalently $|S|<\infty$. The order of a field is the number of its elements.

Theorem 10.1 **Galois**

Finite fields exists for an order q that is a prime power that is $q=p^k$, where $k\geq 1$.

A finite field is also called a Galois field and denoted $GF(q)$.

10.1.7 Divisibility

Definition 10.25

The symbol for divide is $|$. Let $a \in \mathbb{Z}, b \in \mathbb{Z}, a \neq 0$. We write $a|b$, $a \neq 0$, and read a divides b, if there exists an integer $q \in \mathbb{Z}$ such that $b = aq$.

Integer a is then a divisor or factor of b and b is a multiple of a. When a divides b we can also say b is divided (evenly) by a or b is divisible by a, or equivalently the division of b by a leaves a remainder of zero. For $a|b$, a is never a zero; b can be zero.

For a division of b divided by a, integer b is the dividend in such a division. Integer a is the divisor of b also known as a factor of b. The quotient q is $q = b/a = \lfloor b/a \rfloor$, and the integer remainder r is $r = b - a * q$. The remainder is unique if $0 \leq r < |b|$. If $a|b$ then $q = b/a$ is an integer and $r = 0$.

Example 10.1.5

The following integers divide 50.

$$1, 2, 5, 10, 25, 50.$$

We usually write down the positive divisors. The negative divisors are implied.

Definition 10.26 **Trivial divisor**

The trivial divisors of n are 1 and n. If one includes negative numbers, they are 1, -1, n and $-n$.

Definition 10.27

Every integer a divides 0 that is, $\forall a \in \mathbb{Z}$, we have $a|0$.

Definition 10.28

If a does not divide b we write $a \nmid b$ instead.

Definition 10.29 **Odd, Even**

An integer n is even if it is a multiple of two. Otherwise it is an odd integer.

10.1.8 Congruences

Definition 10.30 **Congruence modulo n**

Let $n \in \mathbb{Z}_+$. For $a, b \in \mathbb{Z}$ we say $a \equiv b \pmod{n}$ if the remainder of the division of a by n is equal to the remainder of the division of b by n.

Congruence modulo n and congruence mod n mean the same thing.

Definition 10.31 **Least residue**

If $a \equiv b \pmod{n}$ we refer to b as the residue of a modulo n. For a b such that $0 \leq b < n$, we refer to b as the least non-negative residue of a modulo n.

Note 10.1.1

Oftentimes the least non-negative residue of a modulo n is denoted by $a \bmod n$. Therefore

$$a \equiv b \pmod{n} \Leftrightarrow a \bmod n = b \bmod n.$$

We can read $a \equiv b \pmod{n}$ by saying that "a is congruent to b modulo n". Then the "difference of a and b is a multiple of n" or "n divides the difference of a and b".

Lemma 10.1

Let $n \in \mathbb{Z}_+$, and $a, b \in \mathbb{Z}$. The following conditions are equivalent.

1. $a \equiv b \pmod{n}$,
2. $a = b + nk$, for some $k \in \mathbb{Z}$, and
3. $n|(a-b)$.

Proof.
1. Since $a \equiv b \pmod{n}$, there exist $a_1, b_1, m \in \mathbb{Z}$ such as: $a = a_1 n + m$, $b = b_1 n + m$, where $0 \leq m < n$. Then, $a - b = (a_1 - a_2)n$ i.e. $a = b + (a_1 - a_2)n$ and therefore $a = b + nk$, where $k = (a_1 - a_2)$, thus proving 2. Furthermore, $a - b = (a_1 - a_2)n$ is by definition equivalent to $n|(a-b)$ thus proving 3.
2. Let $a = b + nk$ for some k. Let $b = nq + r$ for some $0 \leq r < n$. The $a = nk + nq + r = n(k+q) + r$. Therefore $a \equiv b \pmod{n}$ thus proving 1. Moreover $a = b + nk$ implies $a - b = nk$ which is equivalent to $n|(a-b)$ thus proving 3.
3. Let $n|(a-b)$. This mean $a - b = nk$ for some integer k. This is 2, and from 2 we might derive 1 as well. ■

Note 10.1.2

The terms congruence and residue, congruence class and residue class will be used interchangeably in the remainder.

Note that the | (divisible/divisibility relation) is reflexive and transitive but not symmetric, and thus it is not an equivalence relation. In fact if $a|b$ and $b|a$ we have $|a| = |b|$ or for positive integers $a = b$.

Lemma 10.2 **Congruence is an equivalence relation**

The congruence $\pmod{n}$ is an equivalence relation as it is reflexive, symmetric and transitive on $\mathbb{Z}$.

Proof.
1. It is straightforward to show $\pmod n$ is reflexive. For every a we have $a \equiv a \pmod n$ since $a-a=0\cdot n$.
2. It is also symmetric since if $a \equiv b \pmod n$ this implies $a-b=nk$ for some integer k. Then $b-a=n(-k)$ and therefore $b \equiv a \pmod n$.
3. Furthermore, $a \equiv b \pmod n$ and $b \equiv c \pmod n$ imply $a-b=kn$ and $b-c=ln$ for some integers k,l. Therefore adding up the equalities we obtain $a-c=(k+l)n$ anf therefore $a \equiv c \pmod n$ thus proving transitivity. ■

Definition 10.32 **Congruence or residue class**

The residue class or congruence class of a mod n or a modulo n is denoted by $(a)_n$.

Definition 10.33 **Congruence or residue classes**

Let $n \in \mathbb{Z}_+$. We define classes $(a)_n$ (or $[a]_n$), for all $a=0,1,\ldots,n-1$, $a \in \mathbb{Z}$.

$$(a)_n = \{b \in \mathbb{Z} | b \equiv a \pmod n\} = \{b = a+kn | k \in \mathbb{Z}\}.$$

$(0)_n, (1)_n, \ldots, (n-1)_n$ are the n equivalence class modulo n. Every integer b belongs to one of those classes depending on its remainder after a division with n. Thus $b \in (a)_n$ if and only if $b \equiv a \pmod n$.

One might use $(a)_n$ or $[a]_n$, or $a+n\mathbb{Z}$, or $a \bmod n$ to denote the equivalence class a modulo (or mod) n. Arithmetic on equivalence classes is known as modular arithmetic.

Corollary 10.1 **mod n**

The $\pmod n$ operation has the following properties.

1. **(Reflexive)** $a \equiv a \pmod n$.
2. **(Symmetric)** $a \equiv b \pmod n \iff b \equiv a \pmod n$.
3. **(Transitive)** If $a \equiv b \pmod n$ and $b \equiv c \pmod n$, then $a \equiv c \pmod n$.
4. **(Translation)** If $a \equiv b \pmod n$ then for any integer c, then $a \pm c \equiv b \pm c \pmod n$.
5. **(Scaling)** If $a \equiv b \pmod n$ then for any integer c, then $a \cdot c \equiv b \cdot c \pmod n$.
6. **(Additivity)** If $a_1 \equiv b_1 \pmod n$ and $a_2 \equiv b_2 \pmod n$, then $a_1+a_2 \equiv b_1+b_2 \pmod n$.
7. **(Subtractivity)** If $a_1 \equiv b_1 \pmod n$ and $a_2 \equiv b_2 \pmod n$, then $a_1-a_2 \equiv b_1-b_2 \pmod n$.
8. **(Multiplicativity)** If $a_1 \equiv b_1 \pmod n$ and $a_2 \equiv b_2 \pmod n$, then $a_1 \cdot a_2 \equiv b_1 \cdot b_2 \pmod n$.
9. **(Exponentiativity)** If $a \equiv b \pmod n$ then for any integer $c \in \mathbb{Z}_+$ then $a^c \equiv b^c \pmod n$.

Proof.
Properties (1)-(3) follow from the equivalence relation proof of congruences mod n.
(4) The translation property follows similarly. If $a \equiv b \pmod n$ then $a=kn+b$. Then $(a+c)=kn+(b+c)$ or $(a-c)=kn+(b-c)$ and the result follows.
(5) The scaling property follows similarly. If $a \equiv b \pmod n$ then $a=kn+b$. Then $(a\cdot c)=(kn+b)\cdot c$ or $ac=(kc)n+bc$ and the result follows.
(6)-(7) They follow similarly to (4)-(5). If $a_1 \equiv b_1 \pmod n$ then $a_1=kn+b_1$. If $a_2 \equiv b_2 \pmod n$ then $a_2=ln+b_2$. Then, $a_1+a_2 \equiv b_1+b_2 \pmod n$, or $a_1-a_2 \equiv b_1-b_2 \pmod n$,
(8) Moreover, $a_1a_2=(kb_2+lb_1+kln)n+b_1b_2$ and thus $a_1 \cdot a_2 \equiv b_1 \cdot b_2 \pmod n$.
(9) For $a \equiv b \pmod n$ implying $a=kn+b$, we use the binomial theorem to show that

$$a^c = (kn+b)^c \equiv b^c \pmod n,$$

as needed. Or it can follow by induction and the scaling property (4). ■

Corollary 10.2

Let $\mathbb{Z}_n$ or $\mathbb{Z}/n\mathbb{Z}$ be the set of equivalence classes of integers modulo n. That is,

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n = \{(0)_n, (1)_n, \ldots, (n-1)_n\}.$$

If $a \in (k)_n$ then $a = nq + k$ i.e. $a \equiv k \pmod{n}$. If $b \in (m)_n$ then $b = nr + m$ i.e. $b \equiv m \pmod{n}$. Then $(a+b) \in (k+m)_n$. Naturally $(k+m)_n$ is $((k+m) \pmod{n})_n$. Moreover $(ab) \in (km)_n$. We can then define operations on the elements of Z_n as follows.

1. **(Addition)** $(k)_n + (m)_n = (k+m)_n$, or $(k+n\mathbb{Z}) + (m+n\mathbb{Z}) = (k+m) + n\mathbb{Z}$.
2. **(Multiplication)** $(k)_n(m)_n = (km)_n$, or $(k+n\mathbb{Z}) \cdot (m+n\mathbb{Z}) = (k \cdot m) + n\mathbb{Z}$.
3. **(Additivity)** If $(a_1)_n = (b_1)_n$ and $(a_2)_n = (b_2)_n$ then $(a_1 + a_2)_n = (b_1 + b_2)_n$.
4. **(Multiplicativity)** If $(a_1)_n = (b_1)_n$ and $(a_2)_n = (b_2)_n$ then $(a_1 \cdot a_2)_n = (b_1 \cdot b_2)_n$.
5. **(Exponentiativity)** If $(a_1)_n = (b_1)_n$ and $c \in \mathbb{Z}_+$, $(a_1^c)_n = (b_1^c)_n$.

Proposition 10.1 **$\mathbb{Z}_n$ or $\mathbb{Z}/n\mathbb{Z}$ is a commutative ring.**

$\mathbb{Z}_n$ or $\mathbb{Z}/n\mathbb{Z}$ is such that $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a commutative ring.

Note 10.1.3

Several times, instead of using the correct notation

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n = \{(0)_n, (1)_n, \ldots, (n-1)_n\},$$

one can use a simplified notation, as follows.

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n = \{0, 1, \ldots, n-1\}.$$

Example 10.1.6

The set of all integers mod 3 fall into three equivalence classes.

$$(0)_3 = \{\ldots, -3, 0, +3, +6, \ldots\},$$

then

$$(1)_3 = \{\ldots, -2, 1, +4, +7, \ldots\},$$

and finally,

$$(2)_3 = \{\ldots, -1, 2, +5, +8, \ldots\}.$$

Integers 5 and 8 belong to $(2)_3$ because the remainder of the integer division of 5 by 3 is a 2. And so is the remainder of the division of 8 by 3.

Corollary 10.3

For a positive integer n,

1. $(a)_n = (b)_n$ if and only if $a \equiv b \pmod{n}$,

2. For $0 \leq a,b < n$ either $(a)_n = (b)_n$ or $(a)_n \neq (b)_n$,

3. There are exaclty n residue class mod n and no more and they cover all of $\mathbb{Z}$.

Proof.
(1) Consider $(a)_n = (b)_n$. Thus for an element x $x \in (a)_n$ if and only if $x \in (b)_n$. If $x \in (a)_n$ then $x = nk + a$. If $x \in (b)_n$ then $x = nl + b$ for some integer k,l. Therefore $nk + a = nl + b$ which leads to $a - b = n(l-k)$ and therefore $a \equiv b \pmod{n}$. Consider $x \in (a)_n$. Then $x \equiv a \pmod{n}$. Since $a \equiv b \pmod{n}$ by transitivity $x \equiv b \pmod{n}$, and thus $x \in (b)_n$. Therefore $(a)_n \subseteq (b)_n$. Similarly we show $(b)_n \subseteq (a)_n$, and the result follows.
(2) Consider a,b. If $a \equiv b \pmod{n}$ then $(a)_n = (b)_n$ from part (1). Otherwise $a \not\equiv \pmod{n}$. Then $(a)_n \neq (b)_n$. If this was not the case, then there would exist $x \in (a)_n \cap (b)_n$. Arguments similar to the proof in (1) then lead to $a \equiv b \pmod{n}$, a contradiction.
(3) Consider integer $a \in \min\{Z\}$ and take the integer division of a by n: $a = nq + r$ where q is the quotient and r the remainder, where $0 \leq r < n$. Then $a \equiv r \pmod{n}$, and thus $a \in (r)_n$. For $0 \leq i, j < n$, $i \neq j$ we have $i \not\equiv j \pmod{n}$. Thus those n congruence classes $((0)_n, \ldots (n-1)_n$ divide $\mathbb{Z}$. ■

Note 10.1.4 **Attention**

In the remainder of this monograph, arithmetic involving $(k)_n$ will not utilize this notation but simply state $k \pmod{n}$ rather than $(k)_n$. Thus we won't use the correct notation $(5)_7 + (3)_7 = (1)_7$. Instead we will write $5 + 3 \equiv 1 \pmod{7}$.

Operation $\pmod{n}$ was used in the context of modular addition, subtraction and multiplication. We were silent about division. This is because of the following example.

$$\begin{array}{rcl} 4 & \not\equiv & 2 \pmod{6} \\ & \text{yet} & \\ 3 \cdot 4 & \equiv & 3 \cdot 2 \pmod{6} \\ & \text{since} & \\ 0 & \equiv & 0 \pmod{6} \\ & \text{Also} & \\ 2 \cdot 3 & \equiv & 0 \pmod{6} \\ & \text{yet} & \end{array}$$

$$2 \not\equiv 0 \pmod{6} \quad \textbf{and} \quad 3 \not\equiv 0 \pmod{6}.$$

Example 10.1.7

Ring $\mathbb{Z}/6\mathbb{Z}$ is not an integral domain. Consider $2 \not\equiv 0 \pmod{6}$. Consider $3 \not\equiv 0 \pmod{6}$. Then $2 \cdot 3 \equiv 0 \pmod{6}$.

Example 10.1.8

$6\mathbb{Z}$ is an ideal I (the set of multiples of 6). Then $\mathbb{Z}/6\mathbb{Z}$ is the set of cosets

$$\{a + I : a \in \mathbb{Z}\},$$

where operations $+, \cdot$ are defined as follows.

$$(a+I) + (b+I) = (a+b) + I,$$

$$(a+I) \cdot (b+I) = (a \cdot b) + I.$$

Two elements a,b are equivalent $a \equiv b$ or equivalently $a + I = b + I$ if $a - b \in I$. If $\mathbb{Z}$ is a ring, $\mathbb{Z}/6\mathbb{Z}$ is a quotient ring also known as a factor ring.

10.1.9 Complete system of congruences

Definition 10.34 **Complete system of congruences**

Let $n \in \mathbb{Z}_+$, and A be a set of integers. A is a complete set of congruences mod n if and only if

- $|A| = n$, and
- A contains exactly one element from each one of the n congruence classes mod n.

Example 10.1.9

Let $n = 3$. Then $A = \{9,4,5\}$ is a complete system of congruences (residues). Note $9 \equiv 0 \pmod 3$, $4 \equiv 1 \pmod 3$, and $5 \equiv 2 \pmod 3$.

Theorem 10.2

Let $n \in \mathbb{Z}_+$, and A be a set of integers. A is a complete set of congruences mod n if and only if

- *$|A| = n$, and*
- *no two elements of A are congruent mod n.*

Proof.
$\Rightarrow$. If A is a complete set of congruences mod n the two conditions above are obviously true.
$\Leftarrow$. If no two elements of A are congruent mod n, then they belong to different congruence classes mod n. Given that A has n elements all the congruence classes are represented in A and thus A is a complete system of congruences. ■

10.1.10 Reduced system of congruences

A congruence class $(b)_n$ is relatively prime to n if and only if $\gcd(b,n) = 1$.

Definition 10.35

Let $n \in \mathbb{Z}_+$.Then $\phi(n)$ is the number of congruence classes modulo n which are relatively prime to n. Then the set of integers

$$\{a_1, a_2, \ldots, a_{\phi(n)}\}$$

defines a reduced system of congruences if the set contains exactly one element from each congruence class mod n that is relatively prime to n.

Definition 10.36 **Reduced system of congruences**

Let $n \in \mathbb{Z}_+$, then $\phi(n)$ is the number of congruence classes modulo n which are relatively prime to n. A congruence class $(b)_n$ is relatively prime to n if and only if $\gcd(b,n) = 1$. Then the set of integers

$$\{a_1, a_2, \ldots, a_{\phi(n)}\}$$

defines a reduced system of congruences if the set contains exactly one element from each congruence class mod n that is relatively prime to n.

Example 10.1.10

Let $n = 10$. Start with

$$\{0,1,2,3,4,5,6,7,8,9\}$$

Delete $0,2,4,5,6,8$ to obtain

$$\{1,3,7,9\}$$

since $0,2,4,5,6,8$ are not relatively prime to 10. This provides a reduced system of congruences Another one is

$$\{11,23,37,-40\},$$

with $11 \in (1)_{10}$ and so on.

Theorem 10.3

Let $n \in \mathbb{Z}_+$. Let A be a set of integers. Then A is a reduced system of congruences mod n if and only if

1. *A has exactly $\phi(n)$ elements,*
2. *no two elements a,b are congruent mod n (i.e. $a \not\equiv b \pmod{n}$ $,\forall a,b \in A$)*
3. *each element of A is relatively prime to n (i.e. $\forall a \in A, \quad \gcd(a,n) = 1$).*

Proof. A reduced set of congruences possesses those properties by default. Let us prove the converse. Let A be a set of integers having those three properties. Because no two elements of A are congruent mod n, it means they belong to difference congruence classes. All classes of A's elements are relatively prime to n. There are $\phi(n)$ elements or classes represented in A. Thus the $\phi(n)$ elements of A represent all $\phi(n)$ classes whose elements are relatively prime to n. Thus A is a reduced set of congruences. ■

Proposition 10.2

If A is a reduced set of congruences, then kA is also one for $\gcd(k,n) = 1$.

10.1.11 Congruence summary

Proposition 10.3

The set $\mathbb{Z}/n\mathbb{Z}$ of integers mod n satisfies the following properties over addition.

1. Closure property: $(i)_n + (j)_n \in \mathbb{Z}/n\mathbb{Z} \quad \forall (i)_n, (j)_n \in \mathbb{Z}/n\mathbb{Z}$.
2. Associativity property: $((i)_n + (j)_n) + (k)_n = (i)_n + ((j)_n + (k)_n) \quad \forall (i)_n, (j)_n, (k)_n \in \mathbb{Z}/n\mathbb{Z}$.
3. Commutativity property: $(i)_n + (j)_n = (j)_n + (i)_n \quad \forall (i)_n, (j)_n \in \mathbb{Z}/n\mathbb{Z}$.
4. Identity that is $(0)_n$, $\forall (i)_n \in \mathbb{Z}/n\mathbb{Z} : (i)_n + (0)_n = (0)_n + (i)_n = (i)_n$.
5. (Additive) inverse property: $\forall (i)_n \in \mathbb{Z}/n\mathbb{Z} \quad \exists (j)_n \in \mathbb{Z}/n\mathbb{Z} : (i)_n + (j)_n = (j)_n + (i)_n = (0)_n$. Moreover $(j)_n = (-i)_n = -(i)_n$.

Proposition 10.4

The set $\mathbb{Z}/n\mathbb{Z}$ of integers mod n satisfies the following properties over multiplication.

1. Closure property: $(i)_n \cdot (j)_n \in \mathbb{Z}/n\mathbb{Z} \quad \forall (i)_n, (j)_n \in \mathbb{Z}/n\mathbb{Z}$.
2. Associativity property: $((i)_n \cdot (j)_n) \cdot (k)_n = (i)_n \cdot ((j)_n \cdot (k)_n) \quad \forall (i)_n, (j)_n, (k)_n \in \mathbb{Z}/n\mathbb{Z}$.
3. Commutativity property: $(i)_n \cdot (j)_n = (j)_n \cdot (i)_n \quad \forall (i)_n, (j)_n \in \mathbb{Z}/n\mathbb{Z}$.
4. Identity that is $(1)_n$, $\forall (i)_n \in \mathbb{Z}/n\mathbb{Z} : (i)_n \cdot (1)_n = (1)_n \cdot (i)_n = (i)_n$.

5. Distribution property of multiplication over addition: $(i)_n \cdot ((j)_n + (k)_n) = ((i)_n \cdot (j)_n) + ((i)_n \cdot (k)_n)$.

Therefore $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a ring, since $(\mathbb{Z}/n\mathbb{Z}, +)$ is an abelian group, $(\mathbb{Z}/n\mathbb{Z}, \cdot)$ is a semigroup, and $\cdot$ is distributive over $+$. Furthermore, $\cdot$ is commutative thus making $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ a commutative ring.

Proposition 10.5

The ring $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a field for a prime integer n.

10.1.12 Units

Let n be a positive integer. The set of all residue classes mod n is denoted by $\mathbb{Z}/n\mathbb{Z}$, and rarely $\mathbb{Z}_n$ or $\mathbb{Z}/n$. It is the set of integers modulo n, thus representing the n equivalence classes that the integers of $\mathbb{Z}$ can be split into depending on the remainder of their division by n. $\mathbb{Z}_n$ is a cyclic group under addition, and a commutative ring under multiplication and addition. The ring is a field for a prime n.

Definition 10.37 **Units in $\mathbb{Z}$**

A divisor of one is called a unit. The only units of $\mathbb{Z}$ are $+1$ and -1.

Definition 10.38 **Alternative definition of a unit in $\mathbb{Z}$**

An element a in $\mathbb{Z}$ that has an inverse in $\mathbb{Z}$ is also known as a unit. Only $+1$ and -1 have inverses in $\mathbb{Z}$.

In $\mathbb{Z}$, $+1$ or -1 are indeed the only invertible elements of $\mathbb{Z}$.

Proposition 10.6

There are no other units in $\mathbb{Z}$, other than $+1$ and -1.

Proof. Let u be a third unit in $\mathbb{Z}$ with $u \neq \pm 1$. Then $1 = uq$ for some integer $q \in \mathbb{Z}$. Since $u \neq \pm 1$ and integer $|u| \geq 2$, noting $1 = uq$ implies $u \neq 0$ as well. But $|u| \geq 2$ and $uq = 1$ implies $|q| \leq 1/2$ and thus $q = 0$. Then $uq = 0$, a contradiction to $uq = 1$. ∎

Consider $\mathbb{Z}/n\mathbb{Z}$ also known as $\mathbb{Z}_n$. Any $g \in \mathbb{Z}/n\mathbb{Z}$ that is invertible and thus $gx \equiv 1 \pmod{n}$ exists, is known as a unit of $\mathbb{Z}_n$ and thus of $\mathbb{Z}/n\mathbb{Z}$. We further conclude that g is a unit if and only if $\gcd(g, n) = 1$. Then $gx \equiv 1 \pmod{n}$ has a solution for $x = g^{-1}$.

Definition 10.39 **Units in $\mathbb{Z}/n\mathbb{Z}$**

Any $g \in \mathbb{Z}/n\mathbb{Z}$ with $\gcd(g, n) = 1$ that is invertible and thus $gx \equiv 1 \pmod{n}$ exists, is known as a unit.

Definition 10.40 **Set of units $\mathbb{U}_n$**

The set of units of $\mathbb{Z}/n\mathbb{Z}$ is denoted by $\mathbb{U}_n$. It is also represented by $(\mathbb{Z}/n\mathbb{Z})^x$, and also represented by $\mathbb{Z}_n^x$.

Definition 10.41

The set of units $\mathbb{U}_n$ has cardinality $\phi(n)$, where $\phi(n)$ is Euler's totient function.

Definition 10.42 **Set of units of ring S is S^x**

Let $(S,+,\cdot)$ be a ring, and let $x \in S$. We say that x is invertible or equivalently that x is a unit, if there exists a $y \in S$ such that

$$x \cdot y = y \cdot x = 1,$$

where the multiplication $\cdot$ implies the multiplicative operation of S, and 1 is the identity element of S over $\cdot$. The set of units of S is denoted by S^x.

Note that for x, the y such that $x \cdot y = y \cdot x = 1$ is uniquely defined. For there was a z such that $x \cdot z = z \cdot x = 1$, then $y = y(xz) = (yx)z = z$. The unique y that is the inverse of x is sometimes denoted as x^{-1}.

Proposition 10.7

For a ring $(S,+,\cdot)$, we have that $(S^x,\cdot)$ is a group.

Let p be a prime. Every non-zero element of $\mathbb{Z}_p$ i.e. $1,\ldots,p-1$ is relatively prime to p, and thus $\phi(p) = p-1$. Moreover $|\mathbb{U}_p| = \phi(p) = p-1$.

Definition 10.43 **Division ring**

Let $(S,+,\cdot)$ be a ring. If all of its non-zero elements are invertible, then S is a division ring. Moreover, if S is commutative then S is a field.

10.1.13 Inverses

Definition 10.44

Let $n \in \mathbb{Z}_+$ and consider $\mathbb{Z}/n\mathbb{Z}$ (also known as $\mathbb{Z}_n$). Let $a \in \mathbb{Z}_n$. The inverse b of a mod n is defined as

$$b \cdot a \equiv 1 \pmod{n}.$$

We can then say that a and b are multiplicative inverses of one another.

We write b as $1/a$ or a^{-1}. Note that implicit in this definition is that $a,b \in \mathbb{Z}/n\mathbb{Z}$ or equivalently $1 \leq a,b < n$. We might use in the remainder for simplicity $\mathbb{Z}_n$ for $\mathbb{Z}/n\mathbb{Z}$.

Proposition 10.8

The inverse of a mod n exists if and only if $\gcd(a,n) = 1$.

This is a byproduct of the extended GCD or results derived for a diophantine equation discussion. We can prove then a more general result.

Theorem 10.4

Consider the ring $(\mathbb{Z}/n\mathbb{Z},+,\cdot)$. The modular equation

$$ax \equiv b \pmod{n},$$

has a solution if and only if $d = \gcd(a,n)$ is such that $d|b$.

Proof.
$\Rightarrow$.
Consider $d = \gcd(a,n)$ thus implying $d|a$ and $d|n$. If $ax \equiv b \pmod{n}$ has a solution, then there exists a $k \in \mathbb{Z}$ such that $ax - b = kn$. By way of $d|a$, $d|n$, we have $d|ax - kn$ and thus $d|b$.

$\Leftarrow$.
For $d = \gcd(a,n)$ it implies $\gcd(a/d, n/d) = 1$. Then a/d is a unit mod n/d. Therefore there exists a y such that

$$(a/d)y \equiv 1 \pmod{(n/d)} \Rightarrow (a/d)y - 1 = k(n/d),$$

for some integer $k \in \mathbb{Z}$. If we multiply this equation by b we derive the following one.

$$(a/d)y - 1 = k(n/d) \Rightarrow a(b/d)y - b = k(b/d)n \Rightarrow a(b/d)y = b + k(b/d)n.$$

Since $d|b$ we have that b/d is an integer and thus we further derive the following.

$$a(b/d)y = b + k(b/d)n \Rightarrow ax = b + Kn,$$

where $x = by/d$ and $K = k(b/d)$. The solution to the modular equation is $x = by/d$, where y is the solution of the modular equation $(a/d)y \equiv 1 \pmod{(n/d)}$. ■

Proposition 10.9

The are $\phi(n)$ numbers a in $\mathbb{Z}_n$ for which the inverse $a^{-1} = 1/a$ exists.

Example 10.1.11

For $n = 10$ there are only $\phi(10) = 4$, a's with inverses. These are $a = 1,3,7,9$ with $a^{-1} = 1,7,3,9$ respectively.

Theorem 10.5

The ring $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a field if and only if n is a prime number.

Proof. From Proposition (10.3) we have that $\mathbb{Z}/n\mathbb{Z}$ is an abelian group over $+$. From Proposition (10.3) we have that $\mathbb{Z}/n\mathbb{Z}$ minus the 0 (additive identity) is an abelian group over $\cdot$, if a (multiplicative) inverse property is true i.e. $\forall (i)_n \in \mathbb{Z}/n\mathbb{Z} \quad \exists (j)_n \in \mathbb{Z}/n\mathbb{Z} : (i)_n \cdot (j)_n = (j)_n \cdot (i)_n = (1)_n$. Moreover $(j)_n = 1/(i)_n = (i)_n^{-1}$. For the latter to be the case n must be a prime number. (This is derived from prior results.) ■

We can then refine the unit definition.

Definition 10.45 **Units in $\mathbb{Z}/n\mathbb{Z}$**

Any $a \in \mathbb{Z}/n\mathbb{Z}$ that has an inverse modulo n, that is, there exists an x that satisfies the following modular equation

$$ax \equiv 1 \pmod{p},$$

is called a unit modulo p. The a is also known as the inverse of a modulo p. Moreover $\gcd(a,n) = 1$.

Lemma 10.3

If $a \in \mathbb{Z}/n\mathbb{Z}$ is a unit modulo n then $\gcd(a,n) = 1$.

10.2 Divisor Properties

Exercise 805.

Uniqueness. Show that for every $a \in \mathbb{Z}^*, b \in \mathbb{Z}$, if $a|b$, there is a unique integer $q \in \mathbb{Z}$ such that $b = aq$.

Solution.
Suppose that $a|b$ i.e. $b = aq$ with $q \in \mathbb{Z}$ and $a \in \mathbb{Z}^*$ that is, $a \neq 0$. If q is not unique, then there might exist a q_1 such that $b = aq_1$ and $q \neq q_1$, since q is not unique.. Then $b = aq = aq_1$ implies $a(q - q_1) = 0$. Since $a \neq 0$, it must be $q - q_1 = 0$. Then $q = q_1$ but this contradicts to the existence of $q_1 \neq q$. ■

Exercise 806.
Show the following for $a \in \mathbb{Z}$.
(a) $a|0$.
(b) $0|0$, and 0 is a divisor of itself and only itself.
(c) $1|a$.
(d) $a|a$.

Solution.
(a) Since $0 = a \cdot 0$ the first claim follows: a is a divisor of 0.
(b) Since $0 = 0 \cdot q$, integer 0 is a divisor of 0. There is no way for any $d \neq 0$ to have $d = 0 \cdot q$. This is because then $d = 0 \cdot q = 0$ which contradicts $d \neq 0$. Thus 0 cannot be the divisor of any $d \neq 0$, $d \in \mathbb{Z}$. Thus 0 only divides 0; moreover, 0 is a multiple of every integer!
(c) $a = 1 \cdot a \Rightarrow 1|a$.
(d) $a = a \cdot 1 \Rightarrow a|a$.

Show that every integer a is a divisor of 0 and that 0 is a divisor of itself and only itself. ■

Exercise 807.
(a) Show the following. Both 5 and -5 are divisors of 5.
(b) Show the following. Both 1 and -1 are divisors of 5.

Solution.
(a) It is $5 = 5 \cdot 1$ and $5 = (-5)(-1)$. Therefore $5, -5, 1, -1$ are divisors of 5. Likewise, $5, -5, 1, -1$ are divisors of -5. We generalize this by replacing 5 with a, and then for every integer a, both $\pm a$ and ± 1 are divisors of a. Therefore a has four divisors as long as $a \neq 0, 1, -1$.
0 has an infinite number of divisors (in fact its set of divisors is $\mathbb{Z}$).
Only $+1, -1$ have two divisors each (itself and the other one of the pair).
$+1$ is the positive unit and -1 the negative unit and collectively are known as the **units** of $\mathbb{Z}$.
(b) See above. ■

Exercise 808.
A number $p \in \mathbb{Z}$ that has only four divisors is called a prime number. Its four divisors are $1, -1, p, -p$. A number $p \in \mathbb{Z}_+$ that has only two positive divisors is called a prime number. Its two such divisors are $1, p$ A number that is not a prime number is a composite number.
(a) Show that 5 is a prime number and 21 is a composite number.

Solution.
$21 = 3 \cdot 7$, and 5 is a prime number by direct inspection (its only two positive divisors are 1 and 5). ■

Exercise 809.
(a) Show that if $a|b$ then $-a|b$, $\quad a|-b$, and finally $-a|-b$.
(b) Show that if $a|b$ and $a|c$, then $a|b\pm c$.
(c) If $a|b$ and $b\neq 0$ then $|a|\leq|b|$.

Solution.
(a) If $a|b$ then there exists $q\in\mathbb{Z}$ such that $b=aq$. Then $b=(-a)(-q)$ and $-b=a(-q)$ and $b=(-a)(-q)$. The three results follow.
(b) We have from $a|b$ that $b=am$ for some $m\in\mathbb{Z}$. Likewise $c=an$ for some $n\in\mathbb{Z}$. Adding or subtracting we have $b\pm c=a(m\pm n)$ and the result follows.
(c) Since $a|b$ we have that $b=am$ for some $m\in\mathbb{Z}$. Since $b\neq 0$ we have $a\neq 0$ as well. Furthermore $m\neq 0$ as well. Thus $|m|\geq 1$. And thus $|b|=|am|=|a||m|\geq|a|$. ■

Exercise 810.

Show that a positive integer a is of the form $a = 2k$ or $a = 2k+1$ for some integer k. Generalize the result for integer a, not just positive integer a.

Solution.
Take the division of a by 2

$$a = 2 \cdot q + r,$$

where q is the integers quotient and r the remainder such that $0 \leq r < 2$. The two possibilities for r are $r = 0$ or $r = 1$. Thus

$$a = 2q, \text{or}, a = 2q + 1.$$

Set $q = k$ and the result follows. The generalization is also obvious. Moreover note that for $a > 0$, $a = 2k$ implies $-a = 2(-k)$. But $a = 2k+1$ implies $-a = 2(-k) - 1 = 2(-k-1) + 1$. ■

Exercise 811.
Show that the square of a positive integer n is of the form $4a$ or $4a+1$, where k is also integer. Generalize the result for integer n.

Solution.
For a (positive) integer n, there are two possibilities for n: n is either an even number or n is an odd number (integer). This depends on the division of n by two. It can have a remainder of 0 or 1. In the former case $n=2k$ and in the latter $n=2k+1$ for some integer k, the quotient of that division. The result then follows, since

$$n=2k \Leftrightarrow n^2=(2k)^2 \Leftrightarrow n^2=4k^2 \Leftrightarrow \quad (a=k^2): \quad n^2=4a,$$

$$n=2k+1 \Leftrightarrow n^2=(2k+1)^2 \Leftrightarrow n^2=4k^2+4k+1 \Leftrightarrow \quad (a=k^2+k): \quad n^2=4a+1.$$

The generalization is also obvious. Start with the generalization of the previous problem. ■

Exercise 812.
Show that if $2|ab$ for $a,b \in \mathbb{Z}$, then $2|a$ OR $2|b$.

Solution.
If neither $2|a$ nor $2|b$, then a,b are odd numbers and thus $a = 2n+1$ and $b = 2m+1$. Then ab must be even. This is not true since

$$ab = (2n+1)(2m+1) = 4mn+2m+2n+1 = 2(2mn+m+n)+1,$$

which is clearly an odd integer number. ■

Exercise 813.

Proposition 10.10
If $a,b,c,d,k,m \in \mathbb{Z}$, then the following apply.

(a) $1|n$. $a|0$ and $0|0$. Moreover, $0|a \Rightarrow a = 0$.

(b1) $a|a$ and of course $\pm a|\pm a$.

(b2) $a|b$ **and** $b|c \Rightarrow a|c$.

(c1) If $a|b$ then for every $k \in \mathbb{Z}$, $a|kb$.

(c2) If $a|b$ then for every $k \in \mathbb{Z}$, $ka|kb$.

(d1) If $a|b$ and $a|c$ then $a|b+c$.

(d2) If $a|b$ and $a|c$ then $a|b-c$.

(d3) If $a|b$ and $a|c$ then $a|kb \pm mc$.

(f1) If $a|b$ and $b \neq 0$ then $|a| \leq |b|$.

(f2) If $a|b$ and $a,b > 0$ then $a \leq b$.

(f3) If $a|b$ and $b|a$ then $|a| = |b|$ or equivalently $a = b = 0$.

(g) If $ka|kb$ and $k \neq 0$ then $a|b$.

Solution.
(a) Previous problems discussed those cases. $n = 1 \cdot n$ implies $1|n$. $0 = a \cdot 0$ implies $a|0$. $0 = 0 \cdot 0$ implies $0|0$, and $0|a$ implies $a = 0 \cdot q$ which by itself implies $a = 0$ and this concludes the case.
(b1) A previous problem discussed this case. $a = a \cdot 1$ implies $a|a$. Moreover $-a = a \cdot (-1)$, and variations conclude the case.
(b2) If $a|b$, then $b = aq$, for some $q \in \mathbb{Z}$. If $b|c$, then $c = br$, for some $r \in \mathbb{Z}$. This implies $c = b \cdot r = (a \cdot q) \cdot r = a \cdot (qr)$ and therefore $a|c$.
(c1) If $a|b$, then $b = aq$, for some $q \in \mathbb{Z}$. Then $kb = kaq = a(kq)$ i.e. $a|kb$.
(c2) If $a|b$, then $b = aq$, for some $q \in \mathbb{Z}$. Then $kb = kaq = (ka)q$ i.e. $ka|kb$.
(d1) If $a|b$, then $b = aq$, for some $q \in \mathbb{Z}$. If $a|c$, then $c = ar$, for some $r \in \mathbb{Z}$. Then $b + c = a(q + r)$ implies $a|b + c$.
(d2) Moreover for the a,b,c of (d1) we have $b - c = a(q - r)$ implies $a|b - c$.
(d3) Furthermore for the a,b,c of (d1) we have $kb \pm mc = a(kq \pm mr)$ implies $a|kb \pm mc$.
(f1) If $a|b$ then $b = aq$. Then $|b| = |aq| = |a| \cdot |q|$. If $b \neq 0$ then $|b| > 0$, since absolute values are positive or zero. This implies that integers a,q are such that $|a| > 0$ and $|q| > 0$. The latter is equivalent to $|q| \geq 1$. Then $|b| = |a| \cdot |q| \geq |a| \cdot 1 \geq |a|$. Equivalently, $|a| \leq |b|$.
(f2) If all of a,b are positive we can drop the absolute values from (f1) concluding $a \leq b$.
(f3) From (f1) we have $|a| \leq |b|$. If $b|a$ we can likewise conclude that $|b| \leq |a|$ thus deriving $|a| = |b|$.
(g) if $ka|kb$ and $k \neq 0$ then $kb = (ka)q$. For non-zero k dividing both sides we have $b = a \cdot q$ and thus $a|b$. ■

Exercise 814.
Find positive integer n such that $n+1|n^2+1$. Then find all such positive integer n.

Solution.
Given that $n^2+1=(n-1)(n+1)+2$, if $n+1|n^2+1$, then $n+1|2$. So, for $n>0$, $n+1=1$ or $n+1=2$. The only positive n value is $n=1$ that satisfies the latter of the two. There are no other n. A more convoluted method works as follows. $n+1|(n+1)^2$. Given that $n+1|n^2+1$ the $n+1|(n+1)^2-n^2-1$ i.e. $n+1|2n$. Since $n+1|2(n+1)$ we conclude $n+1|2(n+1)-2n$ that leads to $n+1|2$ as before. ■

10.3 Compositeness and Primeness

Exercise 815.

Composite and compositeness. An integer $n \in \mathbb{Z}_+{}^*$ is composite if and only if it has a factor a such that $1 < a < n$. (Then there is another factor b such that $1 < b < n$, and $b = n/a$.)

(Positive) Prime number. A positive integer that is not composite is a prime number. A positive integer p that is a prime has factors 1 and p, itself. Alternatively, a prime number p is a positive integer with two factors 1 and itself. Alternatively, a prime number is a positive integer greater than 1 that has only one factor, itself.

Integer Prime number. An integer p is a prime (number) if its possible factors are the units 1 and -1 and the product of itself and the units. Thus the only factors are $1, -1, p, -p$.

Solution.

If n is composite, then n is a multiple of integer a that is neither 1 nor n. Then there exist q such that $n = aq$ for some integer q. If a is neither 1 nor n then $1 < a < n$. (We also used (f2) from Proposition 10.10.) Since $n = aq$, q is also positive and $q > 1$. (If q were equal to one, then $n = a \cdot 1 = n$ i.e. $a = n$ contradicting $1 < a < n$.) Since $a > 1$ $n = aq > q$ implies $q < n$.

A positive integer that is not composite has no factor $1 < a < p$. Therefore its only factors are 1 and p, itself. ■

Exercise 816.

Show the following proposition.

> **Proposition 10.11**
>
> An integer $n \in \mathbb{Z}^*$ with $n > 1$ has a prime factor p i.e. $p|n$.

Solution.

Let Q be the set of natural integer numbers greater than one, that have no prime factors. We shall show that $Q = \emptyset$.

Say Q is not empty. Then there is a minimum element say $m \in Q$. Since $m|m$, and m has no prime factors, m cannot be a prime number. Thus m is composite and let $m = qr$, where $1 < q, r < m$. Since either q or r is $< m$, and m is the smallest element of Q it means $q \notin Q$. By definition q has a prime factor p i.e. $p|q$ and $q|m$. By transitivity $p|m$. The latter contradicts the fact that m being in Q it should not have prime factors (such as p). Thus Q has no minimum element m and thus it is empty! ■

Exercise 817.
For a composite integer $n > 1$ one of its prime divisors p is less than or equal to that is $p \leq \sqrt{n}$ (and such that $p|n$).

Solution.
Let a, b be such that $n = ab$. Let $1 < a \leq b < n$. If $a > \sqrt{n}$ then $b \geq a > \sqrt{n}$ and therefore $n = ab > \sqrt{n} \cdot \sqrt{n} = n$, which is impossible. It follows that $1 < a \leq \sqrt{n}$. Thus a has a prime factor p by Theorem 10.11 and $p \leq a$. Since $p|a$ and $a|n$ we conclude that $p|n$, with $p \leq a \leq \sqrt{n}$. ■

Exercise 818.
Show that $n! \nmid n^n$ for $n > 2$.

Solution.
It suffices to show that $(n-1)! \nmid n^{n-1}$. Assume by way of contradiction that $(n-1)!|n^{n-1}$.
We have two cases two consider.
(a) n is a prime number. Since $(n-1)|(n-1)!$ and $(n-1)!|n^{n-1}$ we have $(n-1)|n^{n-1}$. We have two subcases then.
(a.1) $n-1$ is a prime number. Since $(n-1)|n^{n-1}$ and n is also a prime number (case (a)), we have $n-1|n$. Since $n-1|n-1$ then $n-1|n-(n-1)$, i.e. $n=2$ but this is impossible since $n>2$.
(a.2) $n-1$ is NOT a prime number. From the decomposition theorem, $n-1 = p_1^{a_1} \dots p_k^{a_k}$, where $1 < p_i < n-1$. Since $p_i^{a_i}|(n-1)$ and $(n-1)|n^{n-1}$, we obtain $p_i^{a_i}|n^{n-1}$ i.e. $p_i = n$. This is a contradiction since $p_i < n-1 < n$.
(b) n is NOT a prime number. Consider its decomposition $n = q_1^{b_1} \dots q_t^{b_t}$. Say $n-1 = p_1^{a_1} \dots p_k^{a_k}$ as in case (a.2). Then $p_i|n-1$ and $(n-1)!|n^{n-1}$ i.e. p_i divides one of the q_j. Then $p_i = q_j$ and this is true for all i. Pick any one of them arbitrarily. Say it is p_1. Therefore since $p_1|n-1$ and $p_1 = q_j$ we have that $p_1|n$ i.e. $p_1|(n-(n-1))$ i.e. $p_1|1$ i.e. $p_1 = 1$. This contradicts the primeness of p_1. ■

Exercise 819.

Prime numbers are irreducible. Thus if p is a prime number then p cannot be written in the form $p = qr$ where q, r are both NON units.
Trivial divisors of an integer p are $\pm p$ and ± 1. Non-trivial divisors are called proper divisors. Non-units without proper divisors are called irreducible. A prime number is an irreducible number (in $\mathbb{Z}$).

Solution.
Note 0 cannot be a prime number therefore $p \neq 0$. If a prime number p is NOT irreducible then $p = qr$ where neither q nor r is a unit. Then $p|qr$. Moreover $q|p$ and $r|p$. Then $|q| \leq |p|$ and $|r| \leq |p|$. We claim $p \nmid q$ and $p \nmid r$. If this was not the case (say the former) then $p|q$ i.e. $|p| \leq |q|$ which with $|q| \leq |p|$ implies $q = \pm p$ and then $r = \pm 1$. This contradicts that r is NOT a unit. ■

10.4 A theorem by Euclid

Exercise 820.

Show the following theorem due to Euclid.

Theorem 10.6 **Euclid**

There are infinitely many prime numbers.

Solution.
If there are only finite prime numbers, let ALL of them be $p_1, \ldots, p_m$, where $p_1 = 2$. Then form integer $n = p_1 \cdot p_2 \cdot \ldots \cdot p_m + 1$. There are two possibilities for n: (a) it is a prime number, (b) it is not a prime number.
Case 1. If n is a prime number Then $n = 2 \cdot \ldots \cdot p_i \cdot \ldots \cdot p_m + 1 >> 2p_i + 1 > p_i$ given that all $p_i > 1$. We have just found one more prime number beyond the m ones that were declared ALL that there are: a contradiction.
Case 2. If n is a composite number, let p be a prime factor of n, i.e. $p|n$ with $p > 1$. Such a p exists by a previous problem (composite with a prime factor). This p cannot be one of the m $p_1, \ldots, p_m$. Why ? if p was say $p = p_i$ then $p|p_i$ implies by by (c1),(d3) of a prior problem that $p|p_1 \cdot p_2 \cdot \ldots \cdot p_m$. Since $p|n$ and $n = p_1 \cdot p_2 \cdot \ldots \cdot p_m + 1$ then p divides their difference i.e. $p|1$ by (d2) of a prior problem. This means p is one by (f2) of a prior problem.. But one is a unit not a prime number (plus also the fact that $p > 1$). ■

Exercise 821.

Show that if $n \geq 1$ is an integer then there exists a prime number p such that $n < p \leq n! + 1$.

Solution.
Consider $m = n! + 1$. As previously used, we consider two cases.
Case 1. Integer m is a prime number. We found $p = m$ and we are done.
Case 2. Integer m is NOT a prime number. Then it is a composite and let p be a prime factor. If $p > n$ (and since $p < m$) we are also done and found our p. Otherwise $p \leq n$. Thus $p|n!$ and since p is a factor of m $p|m$ and thus p divides their difference $p|m - n! = p|1$ which leads to p being a unit, a contradiction to its primality. Therefore $p > n$, and its primality is resolved by Case 1 or Case 2 $p > n$ as needed. ■

Exercise 822.
Show that there are no prime numbers between $n!+2$ and $n!+n$.

Solution.
It is clear that $i|n!$ for all $i = 1,2,\ldots,n$. It is clear that $2|n!+2$, and $3|n!+3$, and $4|n!+4$, and therefore $n|n!+n$. Thus all these integers cannot be prime numbers. ■

Exercise 823.
Pick n arbitrary integer numbers between 2 and 10, for any $n \geq 5$. At least two have the same common divisor.

Solution.
The set of prime numbers in the range is 2,3,5,7, i.e. 4 numbers. Let us call this set Y. The set X is the integers from 2 to 10. Define function $f(a) = b$ from X to Y, if integer a has a smallest divisor (other than one course, which is not in Y) that is b. The smallest divisor is to be one of the prime numbers 2,3,5,7. By the choice of X and Y and $|X| > |Y|$ the result follows. Note for example $f(4) = 2$ and $f(6) = 2$.
A simpler way to solve this problem is by inspection! ■

10.5 Integer Division

Exercise 824.

Division. Show the following theorem.

> **Theorem 10.7**
>
> *For $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^*$ there exist unique integers $q \in \mathbb{Z}$ and $r \in \mathbb{Z}$ such that*
>
> $$a = bq + r \quad \textbf{and} \quad 0 \leq r < |b|.$$

Solution.

Case 1 ($a \geq 0, b > 0$). Let A be the set of all non-negative integers $a - bi$, where i is such that $a - bi \geq 0$. For $i = 0$, a belongs to A, and thus A is not empty. By the Well-ordered set principle A has a minimum and let it be r. Since r is in A, we have $r \geq 0$ and $r = a - bi$ for some integer i. All it remains to show is that $r < |b| = b$. Say that $r \geq b$ instead. Then $r - b \geq 0$. Moreover

$$r - b = (a - bi) - b = a - b(i + 1).$$

Thus $r - b \geq 0$ and is of the form $a - bi'$, with $i' = i + 1$. Thus it belongs to A. Moreover $r - b$ is less than r, $r - b < r$ since b is positive. We have found an element smaller than the minimum element of A. This contradicts to the choice of r as being the minimum; we reached contradiction because we assumed $r \geq b$. Thus $r < b$. This thus establishes that $0 \leq r < b$.

We have yet to prove the pair (q, r) is unique. (Note that q is the value of i such that $r = a - bi$.) Let the pair (q, r) be not unique, and let another pair be (q', r'). Then $a = bq + r = bq' + r'$ where $0 \leq r, r' < b$.

This gives

$$b(q - q') = r' - r \Leftrightarrow |b(q - q')| = |r' - r|.$$

Adding $0 \leq r$ and $r' < b$ we get $r' < r + b$ or equivalently $r' - r < b$. Thus $|b(q - q')| < b$. As $b > 0$ we have $|q - q'| < 1$. This can only be possible, if $q = q'$, since q, q' are integer. Then from $a = bq + r = bq' + r'$ if $q = q'$, we deduce also that $r = r'$. This answers positively the question about the uniqueness of the pair (q, r) for case 1.

Case 2 ($a < 0, b > 0$). Similarly as before A is not empty because $a - bi \geq 0$ contains at least one element $a - ba \geq 0$, and the rest of the discussion is similar to case 1.

Case 3 ($a \in \mathbb{Z}, b < 0$). Then $|b| > 0$, and of course $|b| = -b$. By way of cases 1 and 2 for a and $|b|$ we have that $a = |b|q + r$, where $0 \leq r < |b|$. This is equivalent to $a = (-b)q + r$, where $0 \leq r < |b|$. This is equivalent to $a = b(-q) + r$, where $0 \leq r < |b|$. The claim is satisfied for case 3 as well.

Conclusion. For a pair $(a, b \neq 0)$ as defined, there is a unique pair (q, r) with $a = bq + r$ and $0 \leq r < |b|$. ■

Exercise 825.
(a) Find the quotient and the remainder of the division of 37 by 5.
(b) Find the quotient and the remainder of the division of -37 by 5.

Solution.
(a) For 37 we have the following

$$37 = 5 \cdot 7 + 2.$$

37 is the divident, 5 is the divisor and 7 the quotient. The remainder is 2. Note that $0 \leq 2 < |5|$.
(b) For -37 we have the following

$$-37 = 5 \cdot (-8) + 3.$$

-37 is the divident, 5 is the divisor and -8 the quotient. The remainder is 3. Note that $0 \leq 3 < |5|$. One can also observe that

$$-37 = 5 \cdot (-7) - 2.$$

However only the remainder 3 is one that satisfies the $0 \leq 3 < |5|$ requirement and is unique. ■

Exercise 826.

Division related results. For a, b as in division ($b \neq 0$), we have

(i) if $d|a$ and $d|b$ then $d|r$.

(ii) if $d|r$ and $d|b$ then $d|a$.

Solution.
(i) If $d|a, d|b$ then $d|bq$ and thus $d|a-bq$ thus leading (given that $a = bq + r$) to $d|r$.
(ii) If $d|r, d|b$ the $d|bq$ and thus $d|bq+r$ which leads to $d|a$. ■

Exercise 827.

Remainder of division by b. Let $a, A \in \mathbb{Z}$, with $b|a$ and $b|A$. Consider the remainder of the division of a and A with b. The remainders of those two divisions are the same if and only if $a-A$ is a multiple of b.

Solution.
$\Rightarrow$. If $a = bq+r$ and $A = bQ+R$, and we are given $r = R$, the $a-A = bq+r-qQ-r = b(q-Q)$. Since $q-Q \in \mathbb{Z}$, we conclude that $b|a-A$.
$\Leftarrow$. Say that $b|a-A$, i.e. $a-A = bm$ for some $m \in \mathbb{Z}$. Then let $a = bq+r$ and $A = bQ+R$, where $0 \leq r, R < |b|$. From the former $a-A = bm$ and $A = bQ+R$ we have $a-(bQ+R) = bm$ implying $a = b(m+Q)+R$. Because of the uniqueness of (q,r) this latter equality implies $(q,r) = (m+Q, R)$ that is $r = R$ as R is such that $0 \leq R < |b|$).
Result is proven. ■

Exercise 828.
Integer d is such that $d|n$ if and only if the remainder of the division of n by d is 0.

Solution.
Only-if. If d is such that $d|n$ this means $n = qd$ i.e. $n = qd + 0$. Given the uniqueness of q, r from the division problem earlier, we conclude $q = q$ and $r = 0$.
If. If d is such that $n = d \cdot q + r$ with $r = 0$, then $n = d \cdot q$ which implies $d|n$. ■

Exercise 829.

Show the following: if $a^2|b^2$ then $a|b$, for every $a,b \in \mathbb{Z}_+$.

Solution.

There is an easy proof that relies on the prime decomposition of a,b (to be examined later).

Another simple solution that relies on rational numbers vs irrational numbers is as follows. If $a^2|b^2$ then there exists an integer k such that $b^2 = a^2k$. The $a^2/b^2 = k$ and therefore $(b/a) = \sqrt{k}$. Therefore since a,b are integer, then b/a is a rational number. $\sqrt{k}$ is algebraic but also rational; it can't be irrational since that would make b/a irrational contradicting to the rationality of b/a. Since $\sqrt{k}$ is rational, and k integer, there can only be that $\sqrt{k}$ is also integer. Then $b = a\sqrt{k}$ and thus $a|b$.

We can minimally use the well-ordered principle. Let for the sake of contradiction that $a \nmid b$.

$$a^2|b^2 \Rightarrow \exists a_2 : b^2 = a^2a_2. \tag{10.1}$$

Consider,

$$B(a) = \{bt : t \in \mathbb{Z}_+, a^2|b^2 \wedge a|bt\}.$$

Let k be the smallest value such that $bk \in B(a)$. $B(a)$ is not empty. Since $a|b^2$ for $t = b$ we have $bt = b^2 \in B(a)$. The former is true because because $a^2|b^2$ implies $b^2 = a^2r$ and thus $a|a^2r$ implies $a|b^2$ i.e. $a|bt$ for $t = b$. Since k is the smallest t that generates the min element of $B(a)$ by the W.O.S.P. we have by way of $a|bt$, $t = k$.

$$bk = aa_1 \tag{10.2}$$

Consider

$$aba_1 = ab\frac{bk}{a} = b^2k = a^2a_2k. \tag{10.3}$$

We further conclude by dividing by $a \neq 0$, the following

$$aba_1 = a^2a_2k \Rightarrow ba_1 = aa_2k. \tag{10.4}$$

By the division theorem for a_1 and k we have the following.

$$a_1 = kq + r, 0 \leq r < k \Rightarrow r = a_1 - kq \Rightarrow br = ba_1 - bkq \Rightarrow br = aa_2k - aa_1q \Rightarrow br = a(a_2k - a_1q) \Rightarrow a|br \tag{10.5}$$

In the previous derivation we uses in sequence Eq.(10.4), Eq.(10.2). Therefore br is in A since $a^2|b^2$ and $a|br$. Since $r < k$ we have that br is smaller than the smallest element of $B(a)$ which is bk. This leads to contradiction. ■

10.6 Greatest Common Divisor

Greatest Common Divisor. For integers $a, b \in \mathbb{Z}$, $\gcd(a,b)$ is the **greatest common divisor** of a and b if and only if $\gcd(a,b)|a$ and $\gcd(a,b)|b$ and every other divisor c of a,b is such that $c \leq \gcd(a,b)$. Thus

(1) $\gcd(a,b)|a$ and $\gcd(a,b)|b$,

(2) $c|a$ and $c|b \Rightarrow c \leq \gcd(a,b)$.

Moreover, $\gcd(a,b) \leq |a|$ and $\gcd(a,b) \leq |b|$ and of course $\gcd(a,b) > 0$.

Exercise 830.
(a) Show that the $gcd(30,105) = 15$.
(b) Show that the $gcd(5,15) = 5$.
(c) The greatest common divisor of 1 and n is 1.
(d) For $a,b > 0$, if $a|b$ then gcd(a,b)=a.

Solution.
(a) In the remainder we discard negative divisors. For every negative divisor $-a$ there is a positive one a. Thus $-a$ cannot be a gcd, but a can be as $-a < a$ for positive a. The common divisors of 30 and 105 are $\{1,3,5,15\}$. If we include negative numbers then it is $\{\pm 1, \pm 3, \pm 5, \pm 15,\}$, twice as many. The greatest common divisor is indeed 15.
(b) The (positive) divisors of 5 are $\{1,5\}$. The divisors of 15 are $\{1,3,5,15\}$. The (postivie) common divisors are $\{1,5\}$. The greatest common divisor is indeed 5.
(c) The positive divisors of 1 are $\{1\}$. The intersection with the (positive) divisors of n is $\{1\}$ obviously.
(d) Let $d = \gcd(a,b)$. We will show that $d = a$. This will be a result of showing $d \leq a$ and $a \leq d$.
Since $d = \gcd(a,b)$ we conclude $d|a$ and thus $d \leq a$ since $a,b > 0$. Moreovoer $a|a$ and $a|b$ thus $a|\gcd(a,b) = a|d$. Therefore $a \leq d$. The result follows. In the next problem we generalized for arbitrary integers a,b as long as $|a| + |b| \neq 0$ i.e. they are not both zero. ■

Exercise 831.

Let $a,b \in \mathbb{Z}$ not both zero. Then show the following.

(i) $\gcd(a,b) > 0$ and also $\gcd(a,b) \geq 1$.

(ii) $\gcd(a,b) = \gcd(|a|,|b|)$.

(iii) $\gcd(a,b) = \gcd(b,a)$.

(iv) $\gcd(a,1) = 1$.

(v) $\gcd(a,0) = |a|$ for all $a \neq 0$.

(vi) $\gcd(a,b) = |a|$ if and only if $a|b$.

Solution.

Note that if $c|a$ then we have $|c| \leq |a|$. Likewise if $c|b$ we have $|c| \leq |b|$. Combining the two we have $|c| \leq \max(|a|,|b|)$. Thus the set of common divisors of a,b is finite. A finite set of integers always has a maximum, and thus there is a uniquest largest integer $d > 0$ such that $d|a$ and $d|b$. We call d the *greatest common divisor* of a,b and denote it by $\gcd(a,b)$ thus $d = \gcd(a,b)$. Let $S(a)$ be the set of divisors of A. We start showing (i)-(vi).

(i) Obviously. $gcd(a,b)$ is the maximum of the common divisors of a and b i.e. the maximum element of $S(a) \cap S(b)$. One positive element of this set of common divisors is 1 and thus $gcd(a,b) \geq 1$ in addition to being $\gcd(a,b) > 0$.

(ii) Since $S(a) = S(|a|)$ and $S(b) = S(|b|)$ we have that $S(a) \cap S(b) = S(|a|) \cap S(|b|)$. Thus $\gcd(a,b)$ is equal to $\gcd(|a|,|b|)$ since the set of common divisors are equal to each other.

(iii) It is a consequence of the fact that $S(a) \cap S(b) = S(b) \cap S(a)$.

(iv) Since $1|a$ trivially, and the largest divisor of 1 is 1 itself, the result follows. (Note that $S(1) = \{-1,+1\}$.)

(v) $S(a) \cap S(0) = S(a) \cap \mathbb{Z} = S(a)$. The result follows as the largest element of $S(a)$ is $|a|$.

(vi)

$\Rightarrow$. If $\gcd(a,b) = |a|$ then $|a||a$ and $|a||b$. For the latter there exists q such that $b = q|a|$. If a is non-negative, the $b = qa$ as well. If a is negative $b = q(-a) = (-q)a$. The former concludes $a|b$ and so does the latter.

$\Leftarrow$. If $a|b$ then $-a|b$ and thus $|a||b$. Since trivially $a|a$ we conclude that $-a|a$ and thus $|a||a$. Then $|a||\gcd(a,b)$. This by (2) where $c = |a|$, implies $|a| \leq \gcd(a,b)$ as the gcd is always positive and thus no absolute value sign around it is needed. By definition $\gcd(a,b)|a$ and thus $\gcd(a,b)|-a$. Thus $|\gcd(a,b)| \leq |a|$. Combining $|\gcd(a,b)| \leq |a|$ and $|a| \leq \gcd(a,b)$ the result follows. ■

10.7 Extended GCD

Exercise 832.
Extended GCD.

Theorem 10.8 **Bézout identity.**

Let $a,b \in \mathbb{Z}$, where $|a|+|b| \neq 0$. Let $d = \gcd(a,b)$. Show that there exist integers x,y and d such that

$$d = ax + by \tag{10.6}$$

Moreover, show that d is the smallest positive integer that can be written as a linear combination of a,b.

Solution.
Consider

$$A = \{au + bv | u,v \in \mathbb{Z} \text{ and } au + bv > 0\}.$$

Note that for $u = a$, $v = b$, $a \cdot a + b \cdot b > 0$ since $|a| + |b| \neq 0$. Set A has thus at least one element and it is not empty. From the well-ordered set principle there must exist a minimum element for A and let it be d'. Since $d' \in A$ there exist x,y such that

$$ax + by = d'.$$

A. Show d' is a divisor of a. We first show that $d'|a$. Let us form the division operation for the two integers i.e.

$$a = d'q' + r'$$

where $0 \leq r' < d'$. Then we have that

$$r' = a - d'q' = a - (ax + by)q' = a \cdot (1 - xq') + b \cdot (-yq').$$

Note that because of the division of a by d' we know that $0 \leq r'$. We examine two cases.
Case 1: $0 < r'$. If $r' > 0$ then it is a member of A since $r' = a \cdot (1 - xq') + b \cdot (-yq')$. But this cannot happen since $r' < d'$ and d' is the minimum element of A. It would imply the existence of an element of A (i.e. r') smaller than the minimum element of A (i.e. d')!
Case 2: $0 = r'$. The only other possibility is that $r' = 0$. But then, $a = d'q' + r'$ implies $a = d'q'$ i.e. d' is a divisor of a.
B. Show d' is a divisor of b. Similar to the case involving a.
Conclusion. Thus d' divides a and b and is a common divisor of a,b. It should be that $d' \leq d$ since d is the gcd of a and b and is the greatest common divisor of a,b. Furthermore, $d|a$ and $d|b$ and therefore $d|ax+by$ and thus $d|d'$. This implies $d \leq d'$ This $d \leq d'$ along with the $d' \leq d$ shows that $d = d'$. In conclusion

$$d = \gcd(a,b) = d' = ax + by.$$

■

Exercise 833.
Let $a,b \in \mathbb{Z}$, where $|a|+|b| \neq 0$. Let $d = \gcd(a,b)$. Show that set A defined as follows

$$A = \{au+bv | u,v \in \mathbb{Z} \text{ and } au+bv > 0\},$$

contains multiples of d.

Solution.
From the previous problem, the smallest positive element of A was proven to be d. Moreover every element z of A is of the form $z = au+bv$. Since $d|a$ and $d|b$ we have that $d|au+bv$ that is $d|z$. Thus every element of A is a multiple of d. ■

Exercise 834.
Let $a, b \in \mathbb{Z}$, where $|a| + |b| \neq 0$. Let a, b be relatively prime. Show that there exists $x, y \in \mathbb{Z}$ such that

$$ax + by = 1.$$

Solution.
From the previous two problems we know that there exist x, y such that $ax + by = \gcd(a, b)$. Since a, b are relatively prime, the $\gcd(a, b) = 1$ and therefore $ax + by = 1$, as needed. ■

Exercise 835.
Let $a|bc$ and $\gcd(a,b)=1$. Then show that $a|c$.

Solution.
From the previous problem we have that there exist x,y such that $ax+by=1$. Then,

$$acx+bcy=c.$$

Since we are given $a|bc$ we conclude $a|bcy$. Moreover, $a|acx$. Then $a|acx+bcy=a|c$, as needed. ■

Exercise 836.

Show the following proposition

> **Proposition 10.12**
> For a prime number p if $p|p_1p_2$ then $p|p_1$ or $p|p_2$. This can be generalized for $p|p_1p_2p_3\ldots$.

Solution.
If $p|p_1p_2$ given that p is prime its only positive divisors are $1, p$. If $p \nmid p1$, then $\gcd(p, p_1) = 1$. Thus from a prior result $p|p_2$. By induction we can prove its generalization. ■

Exercise 837.
If $\gcd(a,n)=\gcd(b,n)=1$ then show that $\gcd(ab,n)=1$.

Solution.
Let $\gcd(ab,n)=d>1$. Then d has a prime factor and let it be $q>1$ i.e. $q|d$. We have $q|d$, and $d|n$ and therefore $q|n$. Likewise $q|d$ and $d|ab$ and therefore $q|ab$. Since $q|ab$ and q is prime then either $q|a$ or $q|b$. Moreover $q|n$. Then $q|a$ or $q|b$ along with $q|n$ implies $q|\gcd(a,n)$ or $q|\gcd(b,n)$ But $\gcd(a,n)=\gcd(b,n)=1$. Thus $q|1$. This means $q\leq 1$ that contradicts to $q>1$. ■

Exercise 838.

Show that the product of n consecutive integers is divisible by $n!$.
Note. Integer n factorial is defined as follows: $n! = 1 \cdot 2 \cdot \ldots \cdot n$.

Solution.
Let $k+1, k+2, \ldots, k+n$ be n consecutive integers. Their product is $k+1 \cdot k+2 \cdot \ldots \cdot k+n$. We are thus interested in establishing that the following is an integer.

$$\begin{aligned} \frac{k+1 \cdot k+2 \cdot \ldots \cdot k+n}{n!} &= \frac{(k+n)!}{k!n!} \\ &= \binom{k+n}{n}. \end{aligned}$$

This is the number of ways of choosing n distinct objects out of $k+n$ distinct objects. It is an integer! ■

Exercise 839.
If $x, y \in \mathbb{Z}$, then show the following.

$$\gcd(x,y) = \gcd(x, xq+y),$$

for all integers $q \in \mathbb{Z}$.

Solution.
Let $d = \gcd(x,y)$. Let $d_1 = \gcd(x, xq+y)$. If $d|x$ and $d|y$ by the gcd definition, we have $d|xq$ and $d|y$. Thus $d|xq+y$ in addition to $d|x$. Thus d is a common divisor of $xq+y$ and x. Thus $d \leq d_1 = \gcd(x, xq+y)$. For $d_1 = \gcd(x, xq+y)$ we have $d_1|x$ and $d_1|xq+y$. From the former, we conclude that $d_1|xq$; combining it with the latter we have $d_1|xq+y-xq$ i.e. $d_1|y$. Thus $d_1|x$ and $d_1|y$ and therefore d_1 is a common divisor of x, y. Thus $d_1 \leq d = \gcd(x,y)$. By way of $d_1 \leq d$ and $d \leq d_1$ we conclude $d = d_1$.

■

Exercise 840.

In order to compute the $\gcd(a,b)$ we formulate the division operation.

$$a = bq + r,$$

where $0 \leq r < |b|$.
Show then the following

$$\gcd(a,b) = \gcd(bq+r,b) = \gcd(b,r), \quad \text{where} \quad a = bq+r, \quad \text{and} \quad 0 \leq r < |b|. \tag{10.7}$$

Solution.
Let $a = bq + r$, where $0 \leq r < |b|$. Then by way of $\gcd(x,y) = \gcd(x,xq+y)$, we have the following.

$$\gcd(a,b) = \gcd(b,a) = \gcd(b,bq+r) = \gcd(b,r).$$

As a result we have the following

$$\gcd(a,b) = \gcd(bq+r,b) = \gcd(r,b) = \gcd(b,r).$$

■

Exercise 841.
Let $a_1, a_2, \ldots, a_k \in \mathbb{Z}$. Let

$$\begin{aligned} \gcd(a_1, a_2) &= r_1 \\ \gcd(r_1, a_3) &= r_2 \\ \ldots &= \ldots \\ \gcd(r_{n-3}, a_{n-1}) &= r_{n-2} \\ \gcd(r_{n-2}, a_n) &= r_{n-1} \end{aligned} \tag{10.8}$$

Then show that

$$\gcd(a_1, a_2, \ldots, a_n) = r_{n-1}. \tag{10.9}$$

Solution.
We first work bottom to top. From the $n-1$ equations', last equation we have $r_{n-1} | a_n$ and $r_{n-1} | r_{n-2}$. From the penultimate one we have $r_{n-2} | a_{n-1}$ and $r_{n-2} | r_{n-3}$. By transitivity $r_{n-1} | a_{n-1}$ and $r_{n-1} | r_{n-3}$. Therefore we have that r_{n-1} divides all of $a_n, a_{n-1}, \ldots a_2, a_1$ and thus $r_{n-1} \leq \gcd(a_1, a_2, \ldots, a_n)$. We now work top to bottom. Let c be a common divisor of $a_n, a_{n-1}, \ldots a_2, a_1$. Since $c | a_1$ and $c | a_2$ by the first of the sequence of equationswe conclude $c | r_1$. We work similarly to show that c divides all of $r_{n-1}, \ldots r_2, r_1$ and thus $c \leq r_{n-1}$. Therefore any common divisor of $a_n, a_{n-1}, \ldots a_2, a_1$ is such that

$$c \leq r_{n-1} \leq \gcd(a_1, a_2, \ldots, a_n),$$

and one of those common divisors is the gcd implying $\gcd(a_1, a_2, \ldots, a_n) \leq r_{n-1}$. This equation, combined with $r_{n-1} \leq \gcd(a_1, a_2, \ldots, a_n)$ leads to

$$r_{n-1} = \gcd(a_1, a_2, \ldots, a_n),$$

as needed. ■

10.8 Euclid strikes again: GCD algorithm

Exercise 842.

Euclid's algorithm. Show the correctness of the following algorithm for finding the $\gcd(a,b)$.

Theorem 10.9 **Euclid's Algorithm.**

For $a > b$, $b \neq 0$, show that the $\gcd(a,b)$ can be calculated using iterated division, as follows.

$$\begin{array}{lll}
a = bq + r & 0 \leq r < |b| & \gcd(a,b) = \gcd(b,r) \\
b = rq_1 + r_1 & 0 \leq r_1 < r & \gcd(b,r) = \gcd(r,r_1) \\
r = r_1 q_2 + r_2 & 0 \leq r_2 < r_1 & \gcd(r,r_1) = \gcd(r_1,r_2) \\
\quad \ldots & & \\
r_{k-2} = r_{k-1} q_k + r_k & 0 \leq r_k < r_{k-1} & \gcd(r_{k-2},r_{k-1}) = \gcd(r_{k-1},r_k) \\
r_{k-1} = r_k q_{k+1} + 0 & & \gcd(r_{k-1},r_k) = \gcd(r_k,0) = r_k
\end{array}$$

The **last non-zero remainder** *is the gcd:*

$$\gcd(a,b) = r_k.$$

Moreover, for $a > b$ the sequence of steps is finite since $|b| > r > r_1 > r_2 > \ldots > r_{k-1} > r_k > 0$.

Note. The proof of correctness utilizes Eq.(10.7 and in particular $\gcd(a,b) = \gcd(b,r)$.

Solution.

Finite number of steps. The sequence of remainders is a decreasing sequence starting below $|b|$. Thus after no more that $|b|$ steps it will reach a remainder of 0.

Special and trivial cases. If the numbers are negative we consider their absolute values and continue. If $a = b = 0$, there is no $\gcd(a,b)$ as $S(0) = \mathbb{Z}$. If $a > b = 0$, then $\gcd(a,b) = a$. If $b > a = 0$, then $\gcd(a,b) = b$. We deal with the trivial cases next. Otherwise, if $a = b$ then $\gcd(a,b) = a = b$. It is a trivial case.

Top-to-bottom proof: start with $a > b > 0$. If $b > a > 0$, we swap a and b. Thus for all remaining cases we assume that $a > b > 0$ in the remainder.

Therefore, the algorithm computes $\gcd(a,b)$ for $a > b$ and $b \neq 0$. Furthermore, $|a| + |b| \neq 0$. It is obvious from the statement that by using Eq.(10.7 the following apply.

$$\gcd(a,b) = \gcd(b,r) = \gcd(r,r_1) = \gcd(r_1,r_2) = \ldots = \gcd(r_{i-1},r_i) = \ldots = \gcd(r_{k-1},r_k) = \gcd(r_k,0) = r_k.$$

That is, the last, non-zero remainder, is the gcd of a,b. Moreover the sequence of remainders purely decreases i.e. $|b| > r > r_1 > r_2 > \ldots > r_{k-1} > r_k$. Thus after a finite number of no more than b steps r_k will be determined.

Bottom-to-top proof. Starting from the bottom $r_k | r_{k-1}$. Then from the penultimate equation we have r_k dividing both itself and r_{k-1} and thus $r_k | r_{k-1} q_k + r_k$. Thus $r_k | r_{k-2}$. Working likewise we show that r_k divides r and b of the first equation and thus also divides a. Thus r_k divides both a,b. Therefore $r_k \leq \gcd(a,b)$. Pick an arbitrary integer d dividing a and b. Working downwards we show that d divides r_k as well. Thus $d \leq r_k$. Set $d = \gcd(a,b)$. This translates into $\gcd(a,b) \leq r_k$; combined with $r_k \leq \gcd(a,b)$ leads to $r_k = \gcd(a,b)$, as needed. ■

We can show by induction that $r, r_1, \ldots, r_k$ can be written as a linear combination of a,b. For the base case we have $r = a(1) + b(-q)$. Since $r_k = \gcd(a,b)$ this means that

$$r_k = ax + by,$$

for some x,y.

Exercise 843.

(a) Show that in Euclid's algorithm the maximum number of division steps $k+2$ is for a,b that are Fibonacci numbers.
(b) Then show that for $a,b \in \mathbb{Z}$, with $a > b > 0$ the number of divisions in Euclid's Algorithm will be at most $\lg b/\lg\phi + 2 \approx 1.44 \lg b + 2$, where $\phi = (1+\sqrt{5})/2$ is the golden ratio.

Note. $\phi \approx 1.6183$, and $\lg\phi \approx 0.6942$ and $1/\lg\phi \approx 1.44$.

Solution.
The worst-case number of division steps is for two consecutive Fibonacci numbers, say $a = F_n$ and $b = F_{n-1}$, where $F_0 = 0$ and $F_1 = 1$. Then the first division is, $F_n = F_{n-1} + F_{n-2}$, $n > 1$, and the last division $F_2 = F_1 + F_0$, where $F_0 = 0$ and $F_1 = 1$. All quotients generated are one, the smallest possible quotient. The total number of divisions is $n-1$. A bound would be provided for $n-1$ with respect to b.
Note that $F_n \geq \phi^{n-2}$ for all $n \geq 2$. This follows from $\phi^2 = \phi + 1$ and induction on n. Clearly $F_2 = 1 \geq \phi^0$. Moreover

$$F_n = F_{n-1} + F_{n-2} \geq \phi^{n-3} + \phi^{n-4} \geq (1+\phi)\phi^{n-4} \geq \phi^2\phi^{n-4} \geq \phi^{n-2}.$$

If the first division involves $a > b$, this means

$$b \geq F_{n-1} \geq \phi^{n-3} \Leftrightarrow n - 1 \leq \lg b/\lg\phi + 2.$$

Then the number of division steps $n-1$ is upper bounded by $1.44\lg b + 2$. ■

Exercise 844.

Cesaro. For two (uniformly at) random drawn integer a,b, we have the following

$$Pr(\gcd(a,b)=1)\approx 6/\pi^2.$$

Solution.
Consider a prime number p. The chances that an integer n is divisible by p are roughly $1/p$. The probability that $a,b\leq n$ have p as a common factor assuming independence of a,b is roughly $1/p^2$. The probability that they do not is $1-1/p^2$. Let $P=Pr(\gcd(a,b)=1)$. Then

$$P=(1-1/2^2)(1-1/3^2)(1-1/5^2)\ldots$$

and the products includes all prime numbers $2,3,5,7,11,13,\ldots$. Then

$$P=(1+1/2^2+1/2^4+\ldots)^{-1}(1+1/3^2+1/3^4+\ldots)^{-1}(1+1/5^2+1/5^4+\ldots)^{-1}\ldots.$$

Utilizing the Fundamental theorem of arithmetic, every integer is uniquely factorized as a product of two primes. Therefore

$$P=(1+1/2^2+1/3^2+1/4^2+1/5^2+\ldots)^{-1}=\left(\sum_{d=1}^{\infty}\frac{1}{d^2}\right)^{-1}=\left(\frac{\pi^2}{6}\right)^{-1}=\frac{6}{\pi^2}.$$

■

Exercise 845.
Calculate the $\gcd(280,105)$.

Solution.
We note that $280 > 105$ so no swapping is needed.

$$\begin{array}{ll} & : \gcd(280,105) = \\ 280 = 105 \cdot 2 + 70 & : \gcd(105,70) = \\ 105 = 70 \cdot 1 + 35 & : \gcd(70,35) = \\ 70 = 35 \cdot 2 + 0 & : 35 \end{array}$$

Therefore $\gcd(105,280) = 35$. ■

Exercise 846.
For $a = 280$ and $b = 151$, find the $\gcd(a,b)$ and determine x, y such that

$$ax + by = \gcd(a,b).$$

Solution.
We observe that $\gcd(a,b) = \gcd(280,151) = 1$. We use Euclid's algorithm for the gcd as follows.

$$\begin{aligned} 280 &= 151 \cdot 1 + 129 \\ 151 &= 129 \cdot 1 + 22 \\ 129 &= 22 \cdot 5 + 19 \\ 22 &= 19 \cdot 1 + 3 \\ 19 &= 3 \cdot 6 + 1 \\ 3 &= 1 \cdot 3 + 0 \end{aligned}$$

The first non-zero remainder is 1. Thus $\gcd(280,151) = 1$. We then use the extended GCD to find x, y as follows.

$$\begin{aligned} 129 &= 280 - 151 \cdot 1 \\ 22 &= 151 - 129 \cdot 1 \\ 19 &= 129 - 22 \cdot 5 \\ 3 &= 22 - 19 \cdot 1 \\ 1 &= 19 - 3 \cdot 6 \end{aligned}$$

We work from the last equation upwards, one step and one level at a time.

$$\begin{aligned} 1 &= 19 - 3 \cdot 6 \\ &= 19 - (22 - 19 \cdot 1) \cdot 6 = 19 \cdot 7 - 22 \cdot 6 \\ &= (129 - 22 \cdot 5) \cdot 7 - 22 \cdot 6 = 129 \cdot 7 - 22 \cdot 41 \\ &= 129 \cdot 7 - (151 - 129) \cdot 41 = 129 \cdot 48 - 151 \cdot 41 \\ &= (280 - 151) \cdot 48 - 151 \cdot 41 = 280 \cdot 48 - 151 \cdot 89. \end{aligned}$$

Therefore $x = 48$ and $y = -89$.

■

Exercise 847.
Express $\gcd(105,280)$ as a linear combination of 105, 280.

Solution.
$\gcd(105,280) = \gcd(280,105)$. We calculate $\gcd(280,105)$ in a top down form.

$$\begin{aligned} 280 &= 105\cdot 2+70 \\ 105 &= 70\cdot 1+35 \\ 70 &= 35\cdot 2+0 \end{aligned}$$

Therefore $\gcd(280,105) = 35$. We then use the extended GCD to find x,y as follows. We write the equations bottom up starting with the first non-zero remainder that is the gcd.

$$\begin{aligned} 70 &= 280-105\cdot 2 \\ 35 &= 105-70\cdot 1 \end{aligned}$$

We work from the last equation upwards, one step and one level at a time.

$$\begin{aligned} 35 &= 105-70\cdot 1 \\ &= 105-(280-105\cdot 2)\cdot 1 = 280\cdot(-1)+105\cdot 3 \\ &= 280\cdot(-1)+105(3) \end{aligned}$$

Therefore $x=-1$ and $y=3$. ■

Exercise 848.

There is yet another matrix form representation of the Extended-Euclid's algorithms (i.e. Extended GCD). For this in Theorem 10.9 we rewrite the first line as in $a = bq + r = bq_0 + r_0$, and the second line $b = rq_1 + r_1 = r_0q_1 + r_1$. The remaining lines remain the same. The product of $k+2$ matrices can be computed as a 2×2 matrix with entries $x_1, \dots x_4$.

$$\begin{aligned}
\begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} q_0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} b \\ r_0 \end{bmatrix} = \begin{bmatrix} q_0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} q_1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} r_0 \\ r_1 \end{bmatrix} = \dots \\
&= \begin{bmatrix} q_0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} q_1 & 1 \\ 1 & 0 \end{bmatrix} \times \dots \times \begin{bmatrix} q_{k+1} & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} r_k \\ 0 \end{bmatrix} \\
&= \prod_{i=0}^{i=k+1} \begin{bmatrix} q_i & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} r_k \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \times \begin{bmatrix} r_k \\ 0 \end{bmatrix} \\
&\Rightarrow \\
\begin{bmatrix} r_k \\ 0 \end{bmatrix} &= \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}^{-1} \times \begin{bmatrix} a \\ b \end{bmatrix} \\
&\Rightarrow \\
\begin{bmatrix} \gcd(a,b) \\ 0 \end{bmatrix} &= \begin{bmatrix} x & y \\ * & * \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow \gcd(a,b) = ax + by.
\end{aligned}$$

Show this for $a = 1024, b = 640$ and division

$$\begin{aligned}
1024 &= 640 \cdot 1 + 384 \\
640 &= 384 \cdot 1 + 256 \\
384 &= 256 \cdot 1 + 128 \\
256 &= 128 \cdot 2 + 0
\end{aligned}$$

Solution.

$$\begin{bmatrix} r_k \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}^{-1} \times \begin{bmatrix} a \\ b \end{bmatrix}, \qquad \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix}$$

Then,

$$\begin{bmatrix} r_k \\ 0 \end{bmatrix} = \begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix}^{-1} \times \begin{bmatrix} 1024 \\ 640 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ -5 & 8 \end{bmatrix} \times \begin{bmatrix} 1024 \\ 640 \end{bmatrix}$$

We just need to find the top element of the vector i.e. its r_k values. Obviously $r_k = 128 = 1024 \cdot 2 + 640 \cdot (-3)$. At the same time $(x, y) = (2, -3)$ as well. ■

Exercise 849.

Let $a \in \mathbb{N}^*$, $b \in \mathbb{Z}$ such that $a > b$, $b \neq 0$.
Then algorithm ExtendedGCD(a,b) returns (x,y) such that

$$ax + by = \gcd(a,b) = d.$$

Furthermore show that if $d = 1$ and $b > 0$ then $y \bmod a$ and $x \bmod b$ are the inverses of $b \bmod a$ and $a \bmod b$ respectively.

Solution.

Algorithm 1: Extendedgcd(a,b)

Input: a,b, where $a \in \mathbb{N}^*$, $b \in \mathbb{Z}$, $a > b$, $b \neq 0$
Output: (x,y,d)

```
(x,y,d,k,l,m) = (1,0,a,0,1,b) ;
while m > 0 do
    t = d/m;
    (x,y,d,k,l,m) = (k,l,m,x − t·k,y − t·l,d − t·m);
end
return (x,y,d);
```

Details Omitted. ■

Exercise 850.
Trace Extendedgcd(a,b) for $a = 280$ and $b = 105$.

Solution.

$$
\begin{array}{lcll}
\gcd(280,105) & & & \\
(x,y,d,k,l,m) & = & (1,0,280,0,1,105) & , \quad t = 280/105 = 2. \\
(x,y,d,k,l,m) & = & (0,1,105,1-2*0,0-2*1,280-2*105) & , \\
 & = & (0,1,105,1,-2,70) & . \\
(x,y,d,k,l,m) & = & (0,1,105,1,-2,70) & , \quad t = 105/70 = 1. \\
(x,y,d,k,l,m) & = & (1,-2,70,0-1*1,1-1*(-2),105-1*70) & , \\
 & = & (1,-2,70,-1,3,35) & . \\
(x,y,d,k,l,m) & = & (1,-2,70,-1,3,35) & , \quad t = 70/35 = 2. \\
(x,y,d,k,l,m) & = & (-1,3,35,1-2*(-1),-2-2*(3),70-2*35) & , \\
 & = & (-1,3,35,3,-8,0) & . \\
(x,y,d) & = & (-1,3,35) & .
\end{array}
$$

■

Exercise 851.
If $d = \gcd(a,b)$ then $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Solution.
If $d = \gcd(a,b)$ then by Theorem 10.9 we have that $d = r_k$. By the last equation of gcd's we have $\gcd(r_{k-1}, r_k) = r_k = d$. Given that d divides itself ($r_k = d$) we also have that $d|r_{k-1}$. From the prior expression, now that we have $d|r_k$ and $d|r_{k-1}$ using $\gcd(r_{k-2}, r_{k-1}) = \gcd(r_{k-1}, r_k) = r_k = d$ we have that $d|r_{k-2}$ as well. Continuing likewise we have $d|r_2$ and $d|r_1$. Likewise from the third equation of Theorem 10.9 we have $d|r$ and from the second $d|b$. Concluding from the first (division) equation we have $d|a$. We can thus divide by d all equations. This shows that $\gcd(\frac{a}{d}, \frac{b}{d}) = \frac{d}{d} = 1$ as all remainders will be divided by $d = r_k$ as well. ■

Exercise 852.
(a) If $d = \gcd(a,b)$ then $S(a) \cap S(b) = S(d)$. (In other words, $m|a$ and $m|b \iff m|d$.)
(b) Let $a,b \in \mathbb{Z}$, $|a|+|b| \neq 0$. Let $m \in \mathbb{Z}_+^*$. Then

$$\gcd(ma,mb) = m\gcd(a,b).$$

(c) For three $a,b,c \in \mathbb{Z}$, we have $\gcd(a,b,c) = \gcd((a,b),c) = \gcd(d,c)$, where $d = \gcd(a,b)$.

Solution.
(a) Every common divisor of a and b is a divisor of $d = \gcd(a,b)$ since $d = d' = ax+by$. That is $S(a) \cap S(b) \subseteq S(d)$. Moreover for $t \in S(d)$ then $t|d$ and since $d|a$ and $d|b$ by transitivity we have $t|a$ and $t|b$. The former shows that $t \in S(a)$ and the latter that $t \in S(b)$. Both of them show that $t \in S(a) \cap S(b)$. Thus $S(d) \subseteq S(a) \cap S(b)$. This combined with $S(a) \cap S(b) \subseteq S(d)$ leades to the conclusion.
(b) Let $d = \gcd(a,b)$. Since $d|a$ we have $a = dq$ and thus $ma = (md)q$. Thus $md|ma$. Likewise $md|mb$. Thus $md \leq \gcd(mb,ma)$.
Since $m|ma$ and $m|mb$ we have $m|\gcd(ma,mb)$. Thus $\gcd(ma,mb) = mq$ for some q. Moreover $\gcd(ma,mb)|ma$ implies $mq|ma$ i.e $q|a$. Likewise $q|b$. Thus $q|d$ i.e. $q \leq d$. We conclude that $\gcd(ma,mb) = mq \leq md$. From $md \leq \gcd(mb,ma)$ previously and $\gcd(ma,mb) = mq \leq md$ the corollary follows.
(c) $\gcd(a,b,c)$ belongs to $S(a) \cap S(b) \cap S(c) = (S(a) \cap S(b)) \cap S(c)$. The result follows. ■

Exercise 853.
(a) Show that for p a prime, and $a \in \mathbb{Z}^*$, p DOES NOT DIVIDE a if and only if $\gcd(a,p) = 1$.
(b) Show that for $a,b,m \in \mathbb{Z}^*$ and $\gcd(a,b) = 1$ and $a|bm$ then $a|m$.
(c) Show that for $a,b \in \mathbb{Z}^*$ and prime p is such that $p|ab$, then p divides either a or b.
(d) Show that if $\gcd(a,b) = 1$ and $a|c$ and $b|c$ then $ab|c$.

Solution.
(a)
$\Rightarrow$. If p is a prime $S(p) = \{-1,+1,-p,+p\}$, then p does not divide a means that neither p nor $-p$ are divisors of a. The only possible divisors are $+1$ and -1. Thus $S(a) \cap S(p) = \{+1,-1\}$. Thus the gcd is 1, and thus $\gcd(a,p) = 1$ as needed.
$\Leftarrow$. If $\gcd(a,p) = 1$, then p cannot divide a. This is because if $p|a$ since obviously $p|p$ we would have from a prior property that $p|\gcd(a,p)$ and (in fact $\gcd(a,p) = |p|$). For this to happen p should be -1 or $+1$. But p is prime and this can't happen.
(b) Since $\gcd(a,b) = 1$ we have $1 = ax + by$ for some x,y. Multiplying by m we get $m = axm + bmy$. Obviously $a|axm$. Since $a|bm$ we have $a|bmy$ as well. Thus $a|axm + bmy$ i.e. $a|m$.
(c) Say p does not divide a. By (a), since p is prime this means $\gcd(p,a) = \gcd(a,p) = 1$. Since $p|ab$ by the previous theorem we have $p|b$ as prime p must divide one of a,b and it cannot divide a i.e. it must divide b.
(d) If $\gcd(a,b) = 1$ we have $1 = ax + by$. Then $c = acx + bcy$. We have $a|c$ i.e. $c = aq$. Then $cb = abq$ and thus $bcy = ab(qy)$. The latter implies $ab|bcy$.
We also have $b|c$ i.e. $c = br$. Then $ac = abr$. Thus $acx = ab(rx)$. The latter implies $ab|acx$.
Thus $ab|bcy$ and $ab|acx$ imply $ab|acx + bcy$. Therefore $ab|c$. ■

10.9 Least Common Multiple

For $a,b \in \mathbb{Z}$ we denote by $\mathrm{lcm}(a,b)$ the **least common multiple** of a and b, the least positive integer divisible by both a,b. For $a,b \in \mathbb{Z}$ only the positive multiples are then considered.

Exercise 854.
Let $a,b \in \mathbb{Z}$ are such that $|a|+|b| \neq 0$, and let $n = \operatorname{lcm}(a,b)$. If m is a common multiple of a,b show that $n|m$. This says that any common multiple of a,b is a multiple of the least common multiple (of a,b).

Solution.
If one of a,b is equal to zero, then all multiples of zero is zero, common muliples can only be zero and the statement is true trivially.
Otherwise divide m by n.

$$m = nq + r, 0 \leq r < n.$$

We are given that m is a common multiple of a,b that is $a|m$ and $b|m$. Moreover n is the lcm of a,b that is $a|n$ and $b|n$. Therefore $a|r$ and $b|r$. Thus r is a common multiple of a,b. Moreover $r < n$ and the common multiple r is positive and LESS than the least common multiple a,b, a contradiction unless $r = 0$. Then $m = nq$ and thus $n|m$. ■

Exercise 855.
Let $a_1, a_2, \ldots, a_n \in \mathbb{Z}$. Let

$$\begin{aligned} \operatorname{lcm}(a_1, a_2) &= r_1 \\ \operatorname{lcm}(r_1, a_3) &= r_2 \\ \ldots &= \ldots \\ \operatorname{lcm}(r_{n-3}, a_{n-1}) &= r_{n-2} \\ \operatorname{lcm}(r_{n-2}, a_n) &= r_{n-1} \end{aligned} \tag{10.10}$$

Then show that

$$\operatorname{lcm}(a_1, a_2, \ldots, a_n) = r_{n-1}. \tag{10.11}$$

Solution.
We first work top to bottom. It is the case that $r_1|r_2$, $r_2|r_3$, ..., $r_{n-1}|r_n$. It is clear that $a_1|r_1$ and since $r_1|r_2$ we have $a_1|r_2$. Proceeding similarly we have $a_1|r_i$ for all $i = 1, \ldots, n-1$. Likewise we can show $a_i|r_{n-1}$, and thus r_{n-1} is a common multiple of $a_1, \ldots, a_n$ and thus by the previous problem $\operatorname{lcm}(a_1, a_2, \ldots, a_n)$ divides r_{n-1} or equivalently,

$$\operatorname{lcm}(a_1, a_2, \ldots, a_n) \leq r_{n-1}.$$

Let c be a common multiple of $a_1, \ldots a_n$. Then $a_i|c$ for all i. In particular $a_1|c$ and $a_2|c$. Thus c is a common multiple of a_1, a_2. From the previous problem this implies that $\operatorname{lcm}(a_1, a_2)$ which is r_1 divides c. Thus $r_1|c$. From the following equation we have $r_1|c$ just established and $a_3|c$ since c is a common multiple of all a_i. With a similar argument as before we conclude $r_2|c$. By induction we have $r_i|c$ including $r_{n-1}|c$, which generates a

$$r_{n-1} \leq c.$$

Combining the two inequalities derived we have

$$\operatorname{lcm}(a_1, a_2, \ldots, a_n) \leq r_{n-1} \leq c,$$

for every common multiple c of $a_1, \ldots, a_n$, and r_{n-1} is one such common multiple. The only possibility is that $r_{n-1} = \operatorname{lcm}(a_1, a_2, \ldots, a_n)$, as needed, that is there is no common multiple smaller than r_{n-1} because r_{n-1} is the least common multiples of $a_1, \ldots, a_n$. ■

Exercise 856.

(a) Show that $\operatorname{lcm}(am,bm) = m\operatorname{lcm}(a,b)$.
(b) Let $a,b \in \mathbb{N}$. If $\gcd(a,b) = 1$ then show that $\operatorname{lcm}(a,b) = ab$.

Solution.
(a) Let $X = \operatorname{lcm}(a,b)$ and $Y = \operatorname{lcm}(am,bm)$. We are to show that $Y = Xm$. It suffices to show $Y \geq Xm$ and $Y \leq Xm$.
Since $X = \operatorname{lcm}(a,b)$ then a/X and b/X. Therefore $X = ad_1$ and $X = bd_2$ for some d_1, d_2. Consequently $Xm = mad_1$ and $Xm = mbd_2$ respectively. Furthermore $ma|Xm$ and $mb|Xm$. Thefore, since $Y = \operatorname{lcm}(am,bm)$, we have $Xm \geq Y$.
Consider $Y = \operatorname{lcm}(am,bm)$. Them $m|Y$, and therefore there exists a y such that $Y = my$.

$$Y = \operatorname{lcm}(am,bm) \Rightarrow m|Y \Rightarrow \exists y : Y = my.$$

Furthermore by way of $Y = \operatorname{lcm}(am,bm)$ we have $am|Y$ and $bm|Y$ i.e. there exist d_3, d_4 such that $Y = amd_3$ and $Y = bmd_4$.

$$Y = \operatorname{lcm}(am,bm) \Rightarrow am|Y, \quad bm|Y \Rightarrow Y = amd_3, \quad Y = bmd_4.$$

Combining the latter with $Y = my$ we obtain

$$Y = my, Y = amd_3, Y = bmd_4 \Rightarrow my = amd_3, \quad my = bmd_4 \Rightarrow y = ad_3, \quad y = bd_4 \Rightarrow a|y, \quad b|y,$$

$$\Rightarrow a|y, \quad b|y, \Rightarrow y \geq X \Rightarrow my \geq mX \Rightarrow Y \geq Xm.$$

From $Xm \geq Y$ and $Y \geq mX$ we conclude $Y = Xm$ as needed.
(b) Let $m = \operatorname{lcm}(a,b)$. Since $\gcd(a,b) = 1$ we have $1 = ax + by$. Then $m = axm + bym$. Since $a|m$ and $b|m$ we have that $m = ap$ and $m = bq$ for some p,q. Moreover substituting to the previous equation we have $m = ax(bq) + by(ap) = ab(xq) + ab(yp) = ab(xq + yp)$. Thus $ab|m$. This implies $ab \leq m$. Moreover m is the least common multiple of a,b. One such multiple is ab. Thus $ab \geq m$. The $ab \leq m$ and the just shown $ab \geq m$ imply $m = ab$. ■

Exercise 857.
If $T(a)$ is the set of (positive) multiples of a, then $T(a) = T(-a) = T(|a|)$. Moreover $T(a) \cap T(b) = T(|a|) \cap T(|b|)$. Therefore $\operatorname{lcm}(a,b) = \operatorname{lcm}(|a|,|b|)$.
(a) Let $a,b \in \mathbb{N}$. Then show that

$$\gcd(a,b) \cdot \operatorname{lcm}(a,b) = ab.$$

(b) For $a,b \in \mathbb{Z}^*$ show that if $a|b$, then $\operatorname{lcm}(a,b) = b$.

Solution.
(a) Let $d = \gcd(a,b)$. Let $a_1 = a/d$ and $b_1 = b/d$. Then $1 = d/d = \gcd(a/d, b/d) = \gcd(a_1, b_1)$. From the previous problem, part (b), we have that for $\gcd(a/d, b/d) = 1$,

$$\frac{1}{d}\operatorname{lcm}(a,b) = \operatorname{lcm}(a_1,b_1) = \operatorname{lcm}(a/d, b/d) = ab/d^2 = \frac{a \cdot b}{\gcd(a,b)} \Leftrightarrow \operatorname{lcm}(a,b) \cdot \gcd(a,b) = a \cdot b.$$

(b) If $a|b$, then any element $t \in T(b)$ is $t \geq b$. Moreover b or $|b|$ is in $T(a)$. The result follows. ■

10.10 Fundamental Theorem of Arithmetic

Exercise 858.

Fundamental Theorem of Arithmetic. For $n > 1$ there exist unique prime numbers $p_1 < \ldots < p_k$ and natural numbers $a_1, \ldots, a_k > 0$ such that

$$n = p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k}.$$

Solution.
If n is prime this is true; no proof required.
Let n be a composite natural number. Then by Proposition 10.11 it has a prime factor and let its smallest one be q_1 i.e.

$$n = q_1 \cdot a_1, \text{ with } \quad a_1 < n.$$

If a_1 is prime we have decomposed n into two prime numbers. If a_1 is composite then by Theorem 10.11 again, it will have a prime factor and let its smallest one be q_2 i.e.

$$a_1 = q_2 \cdot a_2, \text{ with } \quad a_2 < a_1.$$

$$n = q_1 a_1 = q_1 (q_2 a_2) = q_1 q_2 a_2.$$

If a_2 is composite we repeat this until he hit a $a_{k-1} = q_k$. That way

$$n = q_1 q_2 \ldots q_{k-1} q_k.$$

Let us assume that there is another factorization of n

$$n = r_1 r_2 \ldots r_{t-1} r_t.$$

The we have

$$q_1 q_2 \ldots q_{k-1} q_k = r_1 r_2 \ldots r_{t-1} r_t.$$

r_1 divides the right hand side. It also divides the left-hand side. By Proposition 10.12 one of q_i is divisible by r_1, i.e. $r_1 | q_i$. Because all of r_1, q_i are prime numbers this can only mean $r_1 = q_i$ for some i. The smallest r_i is r_1. The smallest q_i is q_1. Thus $r_1 = q_1$. Because primes are $\neq 0$, we can factor out r_1.

$$q_2 \ldots q_{k-1} q_k = r_2 \ldots r_{t-1} r_t.$$

Continuing likewise if without loss of generality $k < t$ we will eventually have

$$1 = r_{k+1} \ldots r_{t-1} r_t.$$

Prime numbers are > 1. Their product cannot be equal to 1. Thus this can only mean that $k = t$ as well. ■

Exercise 859.
Let $a_i, b_i \in \mathbb{N}$ fo all i. Moreover $a, b \geq 1$.

$$a = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}, \quad b = p_1^{b_1} p_2^{b_2} \dots p_k^{b_k}, \quad a_k, b_k \geq 0.$$

then in order to find $d = \gcd(a,b)$ we have

$$d = \gcd(a,b) = p_1^{c_1} p_2^{c_2} \dots p_k^{c_k} = p_1^{\min(a_1,b_1)} p_2^{\min(a_2,b_2)} \dots p_k^{\min(a_k,b_k)},$$

where $c_i = \min(a_i, b_i) \geq 0$, and then in order to find $m = \operatorname{lcm}(a,b)$ we have

$$m = \operatorname{lcm}(a,b) = p_1^{d_1} p_2^{d_2} \dots p_k^{d_k} = p_1^{\max(a_1,b_1)} p_2^{\max(a_2,b_2)} \dots p_k^{\max(a_k,b_k)},$$

where $d_i = \max(a_i, b_i) \geq 0$,

Solution.
We shall show that $D = p_1^{c_1} p_2^{c_2} \dots p_k^{c_k}$ satisfy the GCD properties. Because $c_i \leq a_i$ for all i, we conclude $D|a$. Because $c_i \leq b_i$ for all i, we conclude $D|b$. Let g by any common divisor of a and b. If g divides a,then $g = p_1^{g_1} p_2^{g_2} \dots p_k^{g_k}$ with $g_i \leq a_i$. If g divides b,then also $g_i \leq b_i$, i.e. $g_i \leq \min(a_i, b_i)$. This means that $g|D$. This is equivalent to also having $g \leq D$. Thus any common divisor g of a, b is $g \leq D$. This means D is $d = \gcd(a,b)$.
We work similarly for the lcm. ■

Exercise 860.
Let $a_i \geq 0, b_i \geq 0$. Moreover $a, b \geq 1$. Let

$$a = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}, \quad b = p_1^{b_1} p_2^{b_2} \dots p_k^{b_k}, \quad a_i, b_i \geq 0, i = 1, \dots, k.$$

Show then

$$\operatorname{lcm}(a,b) \cdot \gcd(a,b) = a \cdot b.$$

Solution.
By the previous problem we have $c_i = \min(a_i, b_i) \geq 0$, and $d_i = \max(a_i, b_i) \geq 0$, and

$$p_i^{c_i} p_i^{d_i} = p_i^{\min(a_i,b_i)} p_i^{\max(a_i,b_i)} = p_i^{a_i} p_i^{b_i}.$$

Therefore

$$\operatorname{lcm}(a,b) \gcd(a,b) = \prod_i p_i^{c_i} p_i^{d_i} = \prod_i p_i^{a_i} p_i^{b_i} = \prod_i p_i^{a_i} \prod_i p_i^{b_i} = a \cdot b.$$

■

Exercise 861.
(a) Find the gcd and lcm of 512, and 4096.
(b) Find the gcd and lcm of 24, and 180.

Solution.
(a) We have $a = 512 = 2^9 = p_1^9$ and $b = 4096 = 2^{12} = p_1^{12}$, where $p_1 = 2$ and $a_1 = 9$ and $b_1 = 12$. Then $c_1 = \min(a_i, b_i) = 9$, and $d_1 = \max(a_i, b_i) = 12$, and $\gcd(a,b) = 2^{c_1} = 2^9 = 512$ and $\operatorname{lcm}(a,b) = 2^{d_1} = 2^{12} = 4096$.
(b) We have

$$a = 24 = 2^3 \cdot 3, \quad b = 180 = 2^2 \cdot 3^2 \cdot 5$$

Then

$$\gcd(a,b) = \gcd(24,180) = 2^{\min(3,2)} 3^{\min(1,2)} 5^{\min(0,1)} = 2^2 3^1 = 12.$$

$$\operatorname{lcm}(a,b) = \operatorname{lcm}(24,180) = 2^{\max(3,2)} 3^{\max(1,2)} 5^{\max(0,1)} = 2^3 3^2 5^1 = 360.$$

■

10.10.1 Mersenne numbers

Exercise 862.

Mersenne numbers. The Mersenne number of order k is denoted by $M_k = 2^k - 1$, where k is a prime number.
Mersenne prime numbers. If $M_k = 2^k - 1$ is a prime number, then M_k is a Mersenne prime (number).
Find the first eight Mersenne numbers. Find the first five Mersenne prime numbers.

Solution.
The first eight Mersenne numbers are as follows.

$$2^2 - 1 = 3, \quad 2^3 - 1 = 7, \quad 2^5 - 1 = 31, \quad 2^7 - 1 = 127,$$

$$2^{11} - 1 = 2047, \quad 2^{13} - 1 = 8191, \quad 2^{17} - 1 = 131071, \quad 2^{19} - 1 = 524287,$$

The first five Mersenne primes are as follows.

$$2^2 - 1 = 3, \quad 2^3 - 1 = 7, \quad 2^5 - 1 = 31, \quad 2^7 - 1 = 127, \quad 2^{13} - 1 = 8191,$$

Note that 2047 is missing from the second list. This is because $2047 = 23 \cdot 89$. One can verify 131071 and 524287 are both Mersenne primes as well. ■

10.10.2 Fermat numbers

Exercise 863.

Fermat numbers. The Fermat number of order k is denoted by $F_k = 2^{2^k} + 1$, $k \geq 0$.
Fermat prime numbers. If F_k is a prime number, then F_k is a prime Fermat number, otherwise it is a composite Fermat number.
(a) Find the first five Fermat numbers. Find the first five prime Fermat numbers.
(b) Show that $F_{n+1} = (F_n - 1)^2 + 1$, and conclude $F_{n+1} - 2 = F_0 F_1 \dots F_n$, or equivalently $F_k | F_{n+1} - 2$, where $0 \leq k \leq n$.

Solution.
(a)

$$F_0 = 3, \quad F_1 = 5, \quad F_2 = 17, \quad F_3 = 257,$$
$$F_4 = 65537, \quad F_5 = 4294967297.$$

All first give are prime Fermat numbers (to mychecking). But for the last one

$$4294967297 = 641 * 6700417,$$

we conclude F_5 is a composite Fermat number.
(b) Consider,

$$(F_n - 1)^2 + 1 = (2^{2^n} + 1 - 1)^2 + 1 = 2^{2 \cdot 2^n} + 1 = F_{n+1}.$$

Moreover

$$F_{n+1} - 2 = 2^{2^{n+1}} - 1 = (2^{2^n} + 1)(2^{2^n} - 1) = F_n(F_n - 2).$$

By induction

$$F_{n+1} - 2 = F_n \cdot (F_n - 2) = F_n \cdot (F_{n-1}(F_{n-1} - 2)) = F_n \cdot F_{n-1}(F_{n-1} - 2) = F_n F_{n-1} \dots F_0.$$

Therefore

$$F_k | F_{n+1} - 2, \quad 0 \leq k \leq n.$$

■

Exercise 864.

There are infinitely many prime numbers distinct from each other.
Note. A different argument is used to prove this result again.

Solution.
Suppose that there are finitely many prime numbers i.e.

$$p_1 < p_2 < \ldots < p_n$$

that is n distinct prime numbers exist. Then form the product $N = p_1 p2 \ldots p_n + 1$. Since $N > 1$ by Theorem 10.11 there is at least one prime p dividing N. This p cannot be any of the $p_1 \ldots p_n$. Why ? Say $p = p_i$ for some i. Then $p|N$ and $p|p_1 \ldots p_n$ which would imply $p|N - p1 \ldots p_n$. The latter implies $p|1$ i.e. $p \leq 1$ but given p is a prime number we must have $p > 1$. A contradiction. Thus p is a prime number other than the ones of the finite group $p_1, \ldots, p_n$. ■

Exercise 865.

A merchant has some pounds of sugar. If the merchant puts them into bags of 11 pounds the merchant can fit enough full bags but then is left with 2 spare pounds. If the merchant uses 20 pound bags then is also left with 1 spare pound. How many pounds of sugar does the merchant have?

Say the merchant has 321 pounds of sugar. This amount needs 29 11-pound bags and there 2 spare pounds. If the merchant uses 20-pound bags the merchant can fill 16 bags and is left with 1 pound. Is it a solution. What if the merchant has 101 pounds of sugar ? $101 = 11 \cdot 9 + 2 = 5 \cdot 20 + 1$.

Solution.

The only solution in $0 \ldots 219$ seems to be 101. But beyond that range, another solution is $101 + 220$, $100 + 2 \cdot 220$, and so on. ■

Exercise 866.
Show that for any positive integer $n > 2$ there is at least one prime number between $> n$ and $< n!$.

Solution.
Consider $N = n! - 1$. If it is a prime number we are done. Question has been proved.
Consider now that $N = n! - 1$ is not a prime number but it is a composite number instead. Then as shown elsewhere, it must have a prime factor and let $p > 1$ be such a prime factor. We shall show that p ish between n and $n!$ i.e. $n \leq p \leq n!$. Say that $p \leq n$. Then $p|n!$ since p appears in $n!$ that is the product of the first n positive integers and p is an integer $\leq n$. Then $p|n!$ and since p is a prime factor of N, we also hve $p|n! - 1$. Then $p|1$ i.e. $p \leq 1$ which contradicts to the fact that p is a prime and thus $p > 1$. Thus p cannot be $p \leq n$. Therefore $p > n$ and of course as factor of $n! - 1$ is $\leq n! - 1 < n!$. ■

10.11 Factorization

Exercise 867.

Unique factorization. Show the following. Let $n \in \mathbb{N}$ with $n > 1$. Then n can be expressed as the product of prime numbers (natural numbers with two divisors one, and itself). The representation is unique up to a permutation of the ordering of the primes.
Note. We usually represent the prime numbers from the smallest (left-most side) to the largest (right-most side). If a prime number p appears k times we also write p^k.
Note. If n itself is a prime number the expression "product" is used degeneratively. The product of one prime is the prime itself.

Solution.
We prove the result by induction: $n > 1$ becomes $n \geq 2$.
Base case $n = 2$. Integer $n = 2$ is the degenerate case of $n = 2$ where there is only one prime factor of $n = 2$ and this is itself.
Induction hypothesis: $H(m)$, $2 \leq m < n$. For $n \geq 2$ we have that for every integer m such that $2 \leq m < n$, then m can be expressed as the products of primes as noted in the statement of the problem.
Inductive step: n Consider integer n. We distinguish two cases: (a) n is a prime number and we are done, (b) n is not a prime number. In the latter case there exists by Proposition 10.11 a prime factor of $1 < p < n$. Call then $P = n/p$. Then $1 < P < n$. By the induction hypothesis P can be expressed as a product of primes, and so does n, since $n = P \cdot p$, where P is a product of primes then, and p is also a prime.
Uniqueness of the representation. Let $p_1 p_2 \dots p_r$ is a representation and $q_1 q_2 \dots q_s$ is another representation of n.

$$n = p_1 p_2 \dots p_{r-1} p_r = q_1 q_2 \dots q_{s-1} q_s,$$

and let $1 < p_1 \leq p_2 \leq \dots \leq p_r$, and $1 < q_1 \leq q_2 \leq \dots \leq q_r$, and furthermore $1 \leq r \leq s$. By using induction we will show that $r = s$ and $p_i = q_i$ for all $i = 1, 2, \dots, r(= s)$. Induction is on r.
Base case $r = 1$. We have

$$n = p_1 = q_1 q_2 \dots q_s.$$

On the left-hand side we have p_1 a prime number. If $s > 1$ on the right-hand side we have the product of at least two primes. This is a contradiction. Therefore $s = 1$. Then $p_1 = q_1$ and we are done.
Induction hypothesis. We assume that the result is true for $r - 1$, where $r \geq 2$. Consider the following.

$$n = p_1 p_2 \dots p_{r-1} p_r = q_1 q_2 \dots q_{s-1} q_s,$$

p_r is a factor of n. It also thus divides the q product. Since p_r is a prime number and all q_i are prime numbers it can't by that $p_r | q_i$ and $p_r < q_i$ because then q_i would have a prime factor and could not be a prime number itself. Thus the only way $p_r | q_i$ is for $p_r = q_i$. In other words $p_r | q_i$ implies $p_r \leq q_i$. We might also work out similarly to show that $q_s \leq p_j$. Then

$$q_s \leq p_j \leq p_r \leq q_i \leq q_s$$

implies $p_r = q_s$. Since both $p_r > 1, q_s > 1$ we divide the expression for n by $p_r = q_s$. We then obtain

$$p_1 p_2 \dots p_{r-1} = q_1 q_2 \dots q_{s-1},$$

and by the induction hypothesis $r - 1 = s - 1$ and $p_1 = q_1$, ..., $p_{r-1} = q_{s-1}$. ■

10.12 Congruences

Exercise 868.

Let $n \in \mathbb{Z}_+^*$. For $a,b \in \mathbb{Z}$ we say $a \equiv b \pmod{n}$ if $n/(a-b)$.

$$\forall a,b \in \mathbb{Z} \quad a \equiv b \pmod{n} \quad \Leftrightarrow \quad n|a-b.$$

We can read this expression by saying that "a is congruent to b modulo n". Then the "difference of a and b is a multiple of n" or "n divides the difference of a and b". Note that several times the $\equiv$ is replaced by $=$ and the parentheses around the mod are dropped.

The $\pmod{n}$ operation has the following properties. Show them.

1. **Reflexive** $a \equiv a \pmod{n}$.
2. **Symmetric** $a \equiv b \pmod{n} \iff b \equiv a \pmod{n}$.
3. **Transitive** If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$.
4. **Translation** If $a \equiv b \pmod{n}$ then for any integer c, then $a+c \equiv b+c \pmod{n}$.
5. **Scaling** If $a \equiv b \pmod{n}$ then for any integer c, then $a \cdot c \equiv b \cdot c \pmod{n}$.
6. **Additivity** If $a_1 \equiv b_1 \pmod{n}$ and $a_2 \equiv b_2 \pmod{n}$, then $a_1+a_2 \equiv b_1+b_2 \pmod{n}$.
7. **Subtractivity** If $a_1 \equiv b_1 \pmod{n}$ and $a_2 \equiv b_2 \pmod{n}$, then $a_1-a_2 \equiv b_1-b_2 \pmod{n}$.
8. **Multiplicativity** If $a_1 \equiv b_1 \pmod{n}$ and $a_2 \equiv b_2 \pmod{n}$, then $a_1 \cdot a_2 \equiv b_1 \cdot b_2 \pmod{n}$.
9. **Exponentiativity** If $a \equiv b \pmod{n}$ then for any integer $c > 0$, then $a^c \equiv b^c \pmod{n}$.

Solution.
1. For $n|a-a$ i.e. $n|0$ obviously.
2. If $n|a-b$ then $n|b-a$ obviously.
3. If $n|a-b$ and $n|b-c$ we have respectively $a-b=kn$ and $b-c=mn$ for some $k,m \in \mathbb{Z}$. Adding the two together we get $a-c=(k+m)n$ i.e. $n|a-c$ i.e. $a \equiv c \pmod{n}$.
4. The $(a+c)-(b+c)=a-b$ proves this property
5. $ca-cb=c(a-b)$ proves this property.
6. The proof is similar to that of property 3.
7. The proof is similar to that of property 3 or 6 except that it involves a subtraction and not addition.
8. If $a_1-b_1=kn$ and $a_2-b_2=mn$, then $a_1a_2=(kn+b_1)(mn+b_2)=b_1b_2+(kmn+kb_2+b_1m)n$, and thus $n|a_1a_2-b_1b_2$.
9. For the last one $n|a-b$ implies $a-b=kn$ or $a=b+kn$. The binomial theorem implies that $a^c=(b+kn)^c=b^c+qn$ for some integer n. The result then follows. ■

Exercise 869.
(a) Let $n \in \mathbb{Z}_+^*$ and $n > 1$ and let $a \in \mathbb{Z}$. Then there exists a unique integer $0 \leq r < n$ such that

$$a \equiv r \pmod{n}.$$

(b) Find $3^{49} \pmod{19}$.

Solution.
(a) If $n \neq 0$ and $a \in \mathbb{Z}$ division implies that there are unique q, r such that $a = nq + r$, with $0 \leq r < |n|$. If in addition n is positive (e.g. $n > 1$), as it is, then $0 \leq r < n$. Then $a - r = nq$ i.e. $a \equiv r \pmod{n}$.
(b) Repeated squares can help avoiding doing 49 multiplications. The binary representation of 49 is $49 = (110001)_2$. Or in other words $49 = 2^5 + 2^4 + 2^0$. Then $3^{49} = 3^{32} \times 3^{16} \times 3^1 = 3^{2^5} \times 3^{2^4} \times 3^{2^0}$.

$$\begin{aligned}
3^1 &\equiv 3 \pmod{19} \\
3^2 &\equiv 9 \pmod{19} \\
3^4 &\equiv 81 \equiv 5 \pmod{19} \\
3^8 &\equiv 25 \equiv 6 \pmod{19} \\
3^{16} &\equiv 36 \equiv 17 \pmod{19} \\
3^{32} &\equiv 289 \equiv 4 \pmod{19}
\end{aligned}$$

Note that $3^{16} \equiv 36 \equiv 17 \equiv -2 \pmod{19}$. Then $3^{32} \equiv 4 \pmod{19}$ and thus we do not need to resort to a calculation of the form $3^{32} = 3^{16} \cdot 3^{16} \equiv 17 \cdot 17 \equiv 4 \pmod{19}$.

We then combine the powers of 2 in the exponent of three as dictacted by the binary representation of 49. That is $3^{49} \equiv 3^1 \pmod{19} \cdot 3^{16} \pmod{19} \cdot 3^{32} \pmod{19} \equiv 3 \cdot 17 \cdot 4 \equiv 13 \cdot 4 \equiv 14 \pmod{19}$

However, a nicer trick might have worked better if a 1 or -1 was encountered earlier.

$$\begin{aligned}
3^1 &\equiv 3 \pmod{19} \\
3^2 &\equiv 9 \pmod{19} \\
3^3 &\equiv 27 \equiv 8 \pmod{19} \\
3^4 &\equiv 24 \equiv 5 \pmod{19} \\
3^5 &\equiv 15 \pmod{19} \\
3^6 &\equiv 45 \equiv 7 \pmod{19} \\
3^7 &\equiv 21 \equiv 2 \pmod{19} \\
3^8 &\equiv 6 \equiv 6 \pmod{19} \\
3^9 &\equiv 18 \equiv -1 \pmod{19}
\end{aligned}$$

The $3^{49} = 3^{45} \cdot 3^4 = (3^9)^5 \cdot 3^4 \equiv (-1)(-1)(-1)(-1)(-1)5 \equiv -5 \equiv 14 \pmod{19}$. ■

Exercise 870.

All integers are assumed to be natural numbers i.e. 0 or positive. Results can be generalized for negative numbers as well.
(a) A single-digit integer is divisible by 5, if it is 0 or 5.
(b) If a two-digit integer ends with 0 or 5 then it is divisible by five. The other way around is also true.
(c) If a three-digit integer ends with 0 or 5 then it is divisible is by five. The other way around is also true.

Solution.
(a) Obviously 0 and 5 are divisible by 5.
(b) Let $n = n_1 n_2$ i.e. $n = 10n_1 + n_2$ If n_2 is 0 or 5 we have the following.

$$10 \equiv 0 \pmod 5 \Rightarrow 10n_1 \equiv 0 \pmod 5,$$

and $0 \equiv 0 \pmod 5$ and $5 \equiv 0 \pmod 5$ and thereofre for $n_2 = 0,5$, we have the following.

$$10n_1 + n_2 \equiv 0 \pmod 5,$$

Since $n = 10n_1 + n_2$ the latter implies

$$n = 10n_1 + n_2 \equiv 0 \pmod 5,$$

i.e. n is a multiple of 5.
(c) Generalize for three digits noting that $10^2 \equiv 0 \pmod 5$. ■

Exercise 871.

If a three-digit integer is divisible by 3, then the sum of its three digits is divisible is by three. The other way around is also true.
Generalize this result to any number of digits.

Solution.
$\Rightarrow$. Let $n = n_1n_2n_3$ be the three digits number. Then $n = 10^2n_1 + 10n_2 + n_3$. We have the following

$$10^2 \equiv 1 \pmod 3 \Rightarrow\ 10^2n_1 \equiv n_1 \pmod 3,$$

$$10 \equiv 1 \pmod 3 \Rightarrow\ 10n_2 \equiv n_2 \pmod 3,$$

and therefore by the additivity rules

$$10^2n_1 + 10n_2 + n_3 \equiv n_1 + n_2 + n_3 \pmod 3.$$

The left-hand side is n. If n is divisible by 3 then $n \equiv 0 \pmod 3$. Therefore,

$$n_1 + n_2 + n_3 \equiv 10^2n_1 + 10n_2 + n_3 \equiv n \equiv 0 \pmod 3,$$

and the result follows.
$\Leftarrow$. All steps are reversible.
Use induction. ■

Exercise 872.

If a three-digit integer is divisible by 9, then the sum of its three digits is divisible is by nine. The other way around is also true.
Generalize this result to any number of digits.

Solution.
Note that $10 \equiv 1 \pmod 9$, $10^2 \equiv 1 \pmod 9$, $10^3 \equiv 1 \pmod 9$, and therefore

$$10^2 n_1 + 10 n_2 + 10 n_3 \equiv n_1 + n_2 + n_3 \pmod 9.$$

Fill-in the details. ■

Exercise 873.

If a three-digit integer is divisible by 11, then the alternating sum (starting from the right) of its three digits is divisible is by eleven.

Solution.
Let $n = n_1 n_2 n_3$ be a 3-digit integer. We are given that

$$n_3 - n_2 + n_1 \equiv 0 \pmod{11}.$$

We will show $n \equiv 0 \pmod{11}$. We note the following.

$$100 \equiv 1 \pmod{11} \Rightarrow 10^2 n_1 \equiv n_1 \pmod{11},$$

and

$$10 \equiv -1 \pmod{11} \Rightarrow 10 n_2 \equiv -n_2 \pmod{11},$$

$$n_3 \equiv n_3 \pmod{11}.$$

Adding the three we obtain the following.

$$n = 10^2 n_1 + 10^1 n_2 + 10^0 n_3 \equiv n_1 - n_2 + n_3 \pmod{11}$$

and since

$$n_3 - n_2 + n_1 \equiv 0 \pmod{11},$$

we conclude

$$n = 10^2 n_1 + 10^1 n_2 + 10^0 n_3 \equiv n_1 - n_2 + n_3 \equiv 0 \pmod{11}.$$

■

Exercise 874.
If $ac \equiv bc \pmod{n}$, and $\gcd(n,c) = 1$ then $a \equiv b \pmod{n}$.

Solution.
It is $ac + qn = bc + rn$ i.e. $ac - bc = sn$ for $s = r - q$ for some integer r, q. Thus $n|ac - bc$ or $n|(a-b)c$. Since $\gcd(n,c) = 1$, we have that $n|a-b$. This implies $a \equiv b \pmod{n}$. ■

10.13 Diophantine equations

Exercise 875.

Diophantine equation. Let $a,b \in \mathbb{Z}^*$. The linear Diophantine equation

$$ax+by=c \tag{10.12}$$

has an integer solution x,y if and only if $d|c$, where $d=\gcd(a,b)$. If a particular solution (x_0,y_0) exists, then there are infinitely many solutions of the form

$$x=x_0+\frac{mb}{d}, \qquad y=y_0-\frac{ma}{d},$$

exist, where $m \in \mathbb{Z}$.

Solution.
$\Rightarrow$. Let x,y be a solution for $ax+by=c$. Then since $d=\gcd(a,b)$ we have $d|a$ and $d|b$ and thus $d|ax+by$ implying $d|c$. If $d \nmid c$ obviously there are no x,y.
$\Leftarrow$. Let $d|c$. By way of the extended GCD algorithm or the definition of $\gcd(a,b)$ there exist x,y such that

$$ax+by=d \tag{10.13}$$

where $d=\gcd(a,b)$. Since $d|c$ there exists a q such that $c=dq$. Multiplying Equation 10.13 by q we get

$$a(xq)+b(yq)=dq=c \tag{10.14}$$

Thus a solution has been found through the definition of gcd or he extended GCD algorithm. The solution is $x_0=xq$ and $y_0=yq$.
Let (x_0,y_0) be a solution pair. Suppose there is another solution pair (x,y). Then

$$ax_0+by_0=c, \quad ax+by=c \tag{10.15}$$

Subtracting one from the other we get

$$a(x_0-x)+b(y_0-y)=0 \Rightarrow a(x_0-x)=b(y-y_0). \tag{10.16}$$

For $d=\gcd(a,b)$, $a_1=a/d$ and $b_1=b/d$ are both integer. Moreover $\gcd(a_1,b_1)=1$ from a prior result. Dividing by d equation 10.16 we have.

$$\frac{a}{d}(x_0-x)=\frac{b}{d}(y-y_0) \Rightarrow a_1(x_0-x)=b_1(y-y_0) \tag{10.17}$$

Since $\gcd(a_1,b_1)=1$, and $a_1|a_1(x_0-x)$ we conclude that $a_1|b_1(y-y_0)$. As a_1,b_1 are relatively prime, this implies $a_1|y-y_0$. Likewise $b_1|x-x_0$. From the latter we obtain that there exists, m such that

$$\begin{aligned} b_1|x-x_0 &\Rightarrow x-x_0=mb_1 \\ &\Rightarrow x=x_0+m\frac{b}{d} \end{aligned}$$

From Eq.(10.16) and $x-x_0=mb_1$ we obtain the following

$$\begin{aligned} x-x_0=mb_1, \quad a_1(x_0-x)=b_1(y-y_0) &\Rightarrow -ma_1b_1=b_1(y-y_0) \\ &\Rightarrow y=y_0-m\frac{a}{d}. \end{aligned}$$

The result thus follows. ■

Exercise 876.

Diophantine equation corollary. Let $a,b \in \mathbb{Z}^*$. The linear Diophantine equation

$$ax+by=c \tag{10.18}$$

has an integer solution x,y if and only if $d|c$, where $d=\gcd(a,b)$. If a particular solution (x_1,y_1) exists for $ax+by=d$, many solutions of the form

$$x=\frac{c}{d}x_1+m\cdot\frac{b}{d}, \qquad y=\frac{c}{d}y_1-m\cdot\frac{a}{d},$$

exist, where $m\in\mathbb{Z}$.

Solution.
Using the extended GCD algorithm, by way of Eq.(10.14), x_1,y_1 are solutions of $ax+by=\gcd(a,b)$. If $d|c$ then there exists a $q\in\mathbb{Z}$ such that $c=dq$. Multiplying Eq.(10.14) by q we derive Eq.(10.15) and thus $x_0=x_1q$ and $y_0=y_1q$ are solutions of Eq.(10.15). Therefore a solution (x_1,y_1) of Eq.(10.14) gives rise to a solution

$$x_0=x_1q=\frac{c}{d}x_1, \quad y_0=y_1q=\frac{c}{d}y_1$$

of Eq.(10.15), and using the previous problem many solutions of the form

$$x=x_0+m\cdot\frac{b}{d}=\frac{c}{d}x_1+m\cdot\frac{b}{d}, \qquad y=y_0-m\cdot\frac{a}{d}=\frac{c}{d}y_1-m\cdot\frac{a}{d},$$

exist, where $m\in\mathbb{Z}$. ■

Exercise 877.
If $\gcd(a,b,c)=1$ and $\gcd(a,b)=d>1$ then Equation 10.12 has no solution.

Solution.
Say that Equation 10.12 has an integer solution (x,y) i.e. $ax+by=c$. Since $d=\gcd(a,b)$ it means $d|a$ and $d|b$. Then $d|ax+by$ i.e. $d|c$. If d is a common divisor of a,b,c it means $d|1$. Then it can only be $d=1$. This contradicts the fact $d>1$. Thus Equation 10.12 cannot have any solution. If $\gcd(a,b)=1$ then Equation 10.12 has one solution. ■

10.13.1 Linear congruences

Exercise 878.

Linear congruences. Show the following. Let $d = \gcd(a,n)$. If $d|b$ the linear congruence below has a solution; if $d \nmid b$ the linear congruence below has NO solutions. The linear congruence

$$ax \equiv b \pmod{n}. \tag{10.19}$$

Solution.
We can rewrite the congruence as a Diophantine equation as follows.

$$\exists k : \ ax - b = kn \Rightarrow ax + n(-k) = b. \tag{10.20}$$

Then, by way of Result (10.12), Eq. (10.19) or its equivalent Eq.(10.20) has a solution if and only if $d|b$, where $d = \gcd(a,n)$. Moreover if one solution is (x_0, y_0) there are more solutions of the form

$$x = x_0 + \frac{mn}{d}, \qquad y = y_0 - \frac{ma}{d},$$

for $m \in \mathbb{Z}$. ■

Exercise 879.

Linear congruences with $d = 1$. Show the following. Let $1 = d = \gcd(a,n)$. If $d|b$ the linear congruence below has one solution.

$$ax \equiv b \pmod{n}. \tag{10.21}$$

Solution.
We can rewrite the congruence as a Diophantine equation as follows.

$$\exists k: \ ax - b = kn \Rightarrow ax + n(-k) = b.$$

If $d = 1 = \gcd(a,n)$ by way of Result (10.12), we have that

$$ax + nk = 1$$

has a solution (x', k') since $\gcd(a,n) = 1$ divides the 1 of the right hand side.

$$ax' + nk' = 1.$$

Then multiplying by b both sides of it we obtain the following.

$$a(bx') + n(bk') = b$$

Therefore the linear congruence

$$ax \equiv b \pmod{n} \Leftrightarrow ax + n(-k) = b,$$

has a solution $x_0 = bx'$, $k_0 = -bk'$.

$$ax_0 + n(-k_0) = b.$$

Other solutions of the linear congruence are

$$x = x_0 + \frac{mn}{d}, \qquad k = k_0 - \frac{ma}{d},$$

for $m \in \mathbb{Z}$. Since $d = 1$ all these are as follows.

$$x = x_0 + mn, \qquad k = k_0 - ma,$$

and thus all other x solutions of the congruence are $x \equiv x_0 \mod n$. Congruence-wise there is only one solution $(x_0)_n$ i.e. x_0. ■

Exercise 880.

Linear congruences with $d > 1$. Show the following. Let $d = \gcd(a,n) > 1$. If $d|b$ the linear congruence below has exactly d solutions.

$$ax \equiv b \pmod{n}. \tag{10.22}$$

Solution.
We can rewrite the congruence as a Diophantine equation as follows.

$$\exists k: \ ax - b = kn \Rightarrow ax + n(-k) = b.$$

If $d = \gcd(a,n) > 1$ by way of Result (10.12), we have that

$$ax + n(-k) = b$$

has a solution (x', k') since $d|b$. Therefore we have the following.

$$ax' + n(-k') = b.$$

Other solutions of the linear congruence are, again from Result (10.12),

$$x = x' + \frac{mn}{d}, \qquad k = k' + \frac{ma}{d},$$

for $m = 0, \ldots, d-1$. This is because

$$a(x' + \frac{mn}{d}) = ax' + \frac{amn}{d} = ax' + nm\frac{a}{d} = ax' + Kn \equiv b \pmod{n},$$

since $\frac{a}{d}$ is an integer by way of $d = \gcd(a,n)$ which implies $d|a$ and thus $K = ma/d$ is also an integer. Note that

$$ax' + n(-k') = b \Rightarrow ax' \equiv b \pmod{n}.$$

Consider the solutions $x_i = x' + (in)/d$ is a $i = 0, \ldots d-1$. Any two solutions x_i, x_j with $0 \leq i < j < d$ are such that $x_i \not\equiv x_j \pmod{n}$. To prove this, say $x_i \equiv x_j \pmod{n}$. Then

$$x_i \equiv x_j \pmod{n} \Rightarrow (i-j)n/d \equiv 0 \pmod{n} \Rightarrow (i-j)n/d = kn \Rightarrow (i-j) = kd,$$

for some integer k. Since $d|kd$ we have $d|i-j$ and thus $d \leq |i-j|$ which is impossible since $|i-j| < d$. Therefore it must be that $x_i \not\equiv x_j \pmod{n}$ for $i \neq j < d$.
Consider now a solution for general $t \geq d$.

$$x_t = x' + (tn)/d.$$

Let $t = Qd + R$, where $0 \leq R < d$. We obtain the following

$$\begin{aligned} x_t &= x' + (tn)/d \\ &= x' + (Qd+R)n/d \\ &= x' + Qn + (Rn)/d \\ &= x' + Qn + (Rn)/d \\ &= x' + (Rn)/d \pmod{n} \end{aligned}$$

implying that any other solution x_t for general t maps modulo n to a solution $x' + (in)/d$ where $i = R$ is in the range $0 \leq i = R < d$.
This concludes the proof that the number of solutions modulo n is indeed $d = \gcd(a,n)$. ■

The latter proof technique (x_i, x_j) is the one used to show Fermat's Little theorem i.e. for a positive integer a with $\gcd(a,p)=1$ we have $a^{p-1} \equiv 1 \pmod{p}$. If we form ia and ja we show that $ia \not\equiv ja \pmod{p}$, for $i \neq j$, and ia for all $i = 1, \ldots, p-1$ form a permutation of $1, 2, \ldots, p-1$ and thus $\prod_{i=1}^{p-1} ia = a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}$. Given that p is prime and $\gcd(i,p)=1$, for all $i = 1, \ldots, p-1$, we conclude $a^{p-1} \equiv 1 \pmod{p}$. Moreover $a^p \equiv a \pmod{p}$.

We can then generalize that for all a we have $a^p \equiv a \pmod{p}$. For a such that $\gcd(a,p)=1$ the result involving Fermat's little theorem is used. If $\gcd(a,p) \neq 1$, given that p is prime implies $p|a$. Then $a^p - a$ is still a multiple of p and therefore $a^p \equiv a \pmod{p}$.

10.13.2 Reduced system of congruences

Exercise 881.

Euler's theorem: reduced system of congruences. Let $n \in \mathbb{Z}_+$ and a such that $\gcd(a,n) = 1$. The following is true.

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

Solution.

If $x_1, \ldots, x_{\phi(n)}$ is a reduced system of congruences mod n, then If $ax_1, \ldots, ax_{\phi(n)}$ is also a reduced system of congruences mod n. Then we have the following

$$\begin{aligned} x_1 \ldots x_{\phi(n)} &\equiv ax_1, \ldots, ax_{\phi(n)} \pmod{n} \\ x_1 \ldots x_{\phi(n)} &\equiv a^{\phi(n)} x_1, \ldots, x_{\phi(n)} \pmod{n} \\ a^{\phi(n)} &\equiv 1 \pmod{n}. \end{aligned}$$

■

Exercise 882.
Let $n \in \mathbb{Z}_+$ and a such that $\gcd(a,n) = 1$. Let $x = a^{-1} \bmod n$. Then

$$x \equiv a^{\phi(n)-1} \pmod{n}.$$

Solution.
Derives from Euler's theorem

$$a^{\phi(n)} \equiv 1 \pmod{n} \Rightarrow aa^{\phi(n)-1} \equiv 1 \pmod{n} \Rightarrow ax \equiv 1 \pmod{n},$$

where $x \equiv a^{\phi(n)-1} \pmod{n}$.

■

Exercise 883.
(a) Solve the modular equation $3x \equiv 11 \pmod 5$.
(b) Solve the modular equation $3x \equiv 5 \pmod{11}$.

Solution.
Note that $\gcd(3,5) = 1$ and $1|11$. Let us find 3^{-1}. We observe $3 \cdot 2 \equiv 1 \pmod 5$ and thus $3^{-1} = 2 \pmod 5$. Then

$$3x \equiv 11 \pmod 5 \Rightarrow x \equiv 3^{-1}11 \pmod 5 \Rightarrow x \equiv 2 \cdot 11 \pmod 5 \Rightarrow x \equiv 2 \pmod 5$$

Note that $\gcd(3,11) = 1$ and $1|5$. Let us find 3^{-1}. We observe $3 \cdot 4 \equiv 1 \pmod{11}$ and thus $3^{-1} = 4 \pmod{11}$. Then

$$3x \equiv 5 \pmod{11} \Rightarrow x \equiv 3^{-1}5 \pmod{11} \Rightarrow x \equiv 4 \cdot 5 \pmod{11} \Rightarrow x \equiv 9 \pmod{11}$$

■

Exercise 884.
Solve $3x \equiv 11 \pmod{14}$.

Solution.
Since $\gcd(3,14) = 1$, $\phi(14) = 6$, we have

$$3^{-1} \equiv 3^{\phi(14)-1} \equiv 3^5 \pmod{14}.$$

We can show $3^5 \equiv 5 \pmod{14}$. The

$$x \equiv 3^{-1}11 \equiv 5 \cdot 11 \pmod{14},$$

which gives $x \equiv 13 \pmod{14}$ i.e. $x = 13$. ■

Exercise 885.
Solve the congruence $6x \equiv 8 \pmod{1}4$.

Solution.
We note that $\gcd(6,14) = 2$ and thus it has two solutions. We divide by $d = 2$ to obtain the following.

$$3x \equiv 4 \pmod{7}.$$

Then $x \equiv 3^{-1} \cdot 4 \equiv 5 \cdot 4 \equiv 6 \pmod{7}$. One solution is $6 \pmod{7}$ i.e. 6 The other solution is $x' + \frac{in}{d} = 6 + \frac{1 \cdot 14}{2} = 13$, $i = 1$. ■

Exercise 886.
Show the following.

$$\gcd(1024, 640) = 128. \tag{10.23}$$

Solution.

$$\begin{aligned} 1024 &= 640 \cdot 1 + 384 \\ 640 &= 384 \cdot 1 + 256 \\ 384 &= 256 \cdot 1 + 128 \\ 256 &= 128 \cdot 2 + 0 \end{aligned}$$

Obviously $\gcd(1024, 640) = 128$, the last non zero remainder. Reversing the order of the equation we have.

$$\begin{aligned} 128 &= 384 + 256 \cdot (-1) \\ &= 384 + (640 + 384 \cdot (-1)) \cdot (-1) \\ &= 640 \cdot (-1) + 384 \cdot (2) \\ &= 640 \cdot (-1) + (1024 + 640 \cdot (-1)) \cdot 2 \\ &= 640 \cdot (-3) + 1024 \cdot 2 \end{aligned}$$

Therefore $\gcd(1024, 640) = 128 = 1024 \cdot 2 + 640 \cdot (-3)$. ■

Exercise 887.
Find the solutions, if any, of $1024x + 640y = 256$.

Solution.
By the previous example $\gcd(1024, 640) = d = 128$. It is $128|256$. Thus one solution of the Diophantine is $x_0 = 2 \cdot (256/128) = 4$ and $y_0 = (-3) \cdot (256/128) = -6$.

Other solutions are

$$x = 4 + m(640/128) = 4 + 5m \quad , \quad y = -6 - 8m$$

A simple calculation confirms the latter solutions

$$1024(4 + 5m) + 640(-6 - 8m) = 256.$$

■

Exercise 888.

The previous method outlined in result 10.23 is tedious. A better approach for solving $ax+by=d$, where $d=\gcd(a,b)$ starts with.

$$\left[\begin{array}{c|cc} a & 1 & 0 \\ b & 0 & 1 \end{array}\right]$$

The first column are the the Dividend (a) and Divisor (b) of the division operation. Eventually through repeated division with remainder (the previous vertical operations become horizontal) will generate a matrix such as the one below. Its first row entries contains the $\gcd(a,b)$ and x,y.

$$\left[\begin{array}{c|cc} \gcd(a,b) & x & y \\ 0 & ? & ? \end{array}\right]$$

Show this for $a=1024, b=640$ and division

$$\begin{aligned} 1024 &= 640\cdot 1+384 \\ 640 &= 384\cdot 1+256 \\ 384 &= 256\cdot 1+128 \\ 256 &= 128\cdot 2+0 \end{aligned}$$

Solution.

$$\begin{aligned} \left[\begin{array}{c|cc} 1024 & 1 & 0 \\ 640 & 0 & 1 \end{array}\right] &\rightarrow \left[\begin{array}{c|cc} 384 & 1 & -1 \\ 640 & 0 & 1 \end{array}\right] \\ &\rightarrow \left[\begin{array}{c|cc} 384 & 1 & -1 \\ 256 & -1 & 2 \end{array}\right] \\ &\rightarrow \left[\begin{array}{c|cc} 128 & 2 & -3 \\ 256 & -1 & 2 \end{array}\right] \\ &\rightarrow \left[\begin{array}{c|cc} 128 & 2 & -3 \\ 0 & -5 & -8 \end{array}\right] \end{aligned}$$

In the first transition, we subtract the second row from the first per the first division. In the second transition, we subtract the first row from the second per the second division. In the third transition, we subtract the second row from the first per the third division. In the fourth transition, we subtract twice the first row from the second per the fourth division. (It is twice because the quotient in this case is a 2.)

■

Exercise 889.

A rational number is an algebraic integer if and only if it is an integer.

Solution.

A number x is an algebraic number if it is the solution of a polynomial with integer coefficients i.e. $f(x)=0$. A integer n is an algebraic number as n is the solution of $f(x)=0$ for $f(x)=x-n$. Likewise a rational a/b, $b\neq 0$, is an algebraic number as a/b is the solution of $f(x)=0$ for $f(x)=bx-a$. If a number is not algebraic number then it is a transcedental number. A number n is an algebraic integer if it is the root of a non-zero monic polynomial with integer coefficients (monic: coefficient of one for leading degree defining term).

An integer is always and algebraic integer as $f(x)=x-n$ is a monic polynomial. Let a/b is be an algebraic integer. We will show that it is an integer, in other words $b=1$. Let $q=a/b$ is in its simplest form $\gcd(a,b)=1$. Then q as an algebraic integer it is the solution of say

$$f(x)=x^n+f_{n-1}x^{n-1}+\ldots+f_1x+f_0.$$

Since $f(a/b)=0$, substituting a/b for x above we obtain the following.

$$0=a^n+f_{n-1}a^{n-1}b+\ldots+f_1ab^{n-1}+f_0b^n.$$

We have that b divides all but the first term of the right hand side. Thus $b|a^n$. Since $\gcd ba=1$ this implies $b=1$ as otherwise b,a have a common divisor and then $\gcd(a,b)\neq 1$ contradicting $\gcd(a,b)=1$. The result follows. ■

10.14 Short review on primitive roots and quadratic residues

10.14.1 Euler's totient function

Definition 10.46

For any $n \in \mathbb{N}$, let

$$n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k},$$

where $p_1, p_2, \dots, p_k$ are k distinct prime numbers, $k \geq 1$, and $a_1, a_2, \dots a_k \in \mathbb{N}$ with $a_i \geq 1$. We define $\phi(n)$ as follows.

$$\phi(n) = p_1^{a_1-1}(p_1-1)p_2^{a_2-1}(p_2-1)\dots p_k^{a_k-1}(p_k-1).$$

Moreover if we define $\phi(1) = 1$ then we can write $\phi(n)$ as follows as well.

$$\phi(n) = \left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\dots\left(1-\frac{1}{p_k}\right).$$

Definition 10.47

Function $\phi(n)$ denotes the cardinality of $\mathbb{Z}_n$ i.e. $\phi(n) = |\mathbb{Z}_n|$, the number of units $\pmod n$.

Corollary 10.4

If p, q are prime numbers and then $\gcd(p,q) = 1$, then

$$\phi(pq) = \phi(p)\phi(q).$$

Proof. The number of integers relatively prime to pq are the multiples of p, the multiple of q, and if we include in this count pq make sure it is not included twice. Thus out of the pq integers $1, 2, \dots, pq$. There are $pq/p = q$ multiples of p (including pq) and there are $pq/q = p$ multiples of q (including pq), and thus the number of integers relatively prive to pq are

$$pq - p - q + 1,$$

where the $+1$ is because in $-p$ and $-q$ the contribution of pq was subracted twice and we adjust with a $+1$ accordingly. Then

$$pq - p - q + 1 = (p-1)(q-1) = \phi(p)\phi(q).$$

■

Corollary 10.5

If p, q are are relatively prime and then $\gcd(p,q) = 1$, then

$$\phi(pq) = \phi(p)\phi(q).$$

Proof. See one of the problems to follow. ■

10.14.2 Units, zero divisors, quadratic residues

Any $a \in \mathbb{Z}_n$ that is invertible and thus $ax \equiv 1 \pmod n$ exists, is known as a unit. We further conclude that a is a unit if and only if $\gcd(a,n) = 1$. Then $ax \equiv 1 \pmod n$ has a solution for $x = a^{-1}$.

Definition 10.48 **Units in $\mathbb{Z}_n$**

Any $a \in \mathbb{Z}_n$ that has an inverse modulo n, that is, there exists an x that satisfies the following modular equation

$$ax \equiv 1 \pmod{n}$$

is called a unit modulo n. The x is also known as the inverse of a modulo n.

Definition 10.49 **Zero divisors in $\mathbb{Z}_n$**

An a that is not a unit is called a zero divisor modulo n.

Definition 10.50 **Set of units $\mathbb{U}_n$**

The set of units of $\mathbb{Z}_n$ is denoted by $\mathbb{U}_n$. It is also represented by $\mathbb{Z}_n^x$. It is also represented by $(\mathbb{Z}/n\mathbb{Z})^x$.

Definition 10.51 **Cardinality of $\mathbb{U}_n$**

The set of units $\mathbb{U}_n$ has cardinality $\phi(n)$, where $\phi(n)$ is Euler's totient function.

Let p be a prime. Every non-zero element of $\mathbb{Z}_p$ i.e. $1, \ldots, p-1$ is relatively prime to p, and thus $\phi(p) = p-1$. Moreover $|\mathbb{U}_p| = \phi(p) = p-1$.

Definition 10.52 **Quadratic residue**

For $p \in \mathbb{N}$, where p is an odd prime number and for any $a \in \mathbb{Z}$ such that $p \nmid a$ we say that a is a quadratic residue $\pmod{p}$ if and only if the congruence

$$x^2 \equiv a \pmod{p}$$

has a solution.

The definition extends to the general case of an integer n rather than of odd prime p.

Definition 10.53 **Quadratic residue**

For $n \in \mathbb{N}$, and for any $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$, we say that a is a quadratic residue $\pmod{n}$ if and only if the congruence

$$x^2 \equiv a \pmod{n}$$

has a solution.

Let $a \equiv b \pmod{p}$, for p an odd prime number, then $x^2 \equiv b \pmod{p}$ has a solution if and only if $x^2 \equiv a \pmod{p}$ does also. Moreover $p|b$ if and only if $p|a$. Quadratic residuocity (with respect to odd prime number p) is relevant to the introduction of the Legendre symbol. Later we extend it to the introduction of the Jacobi symbol.

10.14.3 Legendre symbol

The Legendre symbol $\left(\frac{a}{p}\right)$ is defined to be 1 if a is a quadratic residue $\pmod{p}$ and -1 if it is a quadratic non-residue (i.e. it is not a quadratic residue mod p). It is 0 if $p|a$.

Definition 10.54 **Legendre symbol**

Let $p \in \mathbb{N}$ be an odd prime number. For any $a \in \mathbb{Z}$ the Legendre symbol is defined as follows.

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if} p|a \\ 1 & \text{if a is quadratic residue mod p} \\ -1 & \text{otherwise.} \end{cases}$$

Note that implicit in the two latter cases is the fact that $p \nmid a$.
Viewing the Legendre symbol as a function we conclude the following.

$$\left(\frac{a+kp}{p}\right) = \left(\frac{a}{p}\right).$$

10.14.4 Order of a mod n

$\mathbb{Z}$ is the set of integers (positive, negative or zero). For a positive integer $n > 0$ the set $\mathbb{Z}_n$ sometimes denoted by $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}/n$ is the set of integers modulo n, thus representing the n equivalence classes that the integers of $\mathbb{Z}$ can be split into depending on the remainder of their division by n.

$\mathbb{Z}_n$ is a cyclic group under addition, and a commutative ring under multiplication and addition. The ring is a field for a prime n.

Definition 10.55 **Order mod n**

For $n \in \mathbb{Z}$, $n > 0$, and $a \in \mathbb{Z}_n$ with $\gcd(a,n) = 1$ we define $\mathrm{ord}_n(a) = k$, the order of a mod n, to be the smallest positive integer k such that $a^k \equiv 1 \pmod{n}$.

By Euler's theorem $a^{\phi(n)} \equiv 1 \pmod{n}$ implies that $k|\phi(n)$. (If this was not the case $\phi(n) = kq + r$, $0 \leq r < k$ would imply $a^r \equiv 1 \pmod{n}$ for an $r < k$, an impossibility given that $\mathrm{ord}_n(a) = k$.) An element a as defined above that has order $\phi(n)$ is called a primitive root mod n.

10.14.5 Primitive roots mod n

Definition 10.56 **Primitive root in $\mathbb{Z}$**

Let $n \in \mathbb{Z}_+$. An integer $g \in \mathbb{Z}$ is a primitive root modulo n if $g \bmod n \in \mathbb{Z}_n^x = U_n^x$ has order $\phi(n)$.

Definition 10.57 **Primitive roots in $\mathbb{Z}_n^x$**

A unit $g \pmod{n}$, where $g \in \mathbb{Z}_n$ and $\gcd(g,n) = 1$, is a primitive root mod n, if its order is $\phi(n)$.

Corollary 10.6

A unit $g \pmod{n}$ is a primitive root if it generates all $\mathbb{U}_n$.

Proof. That is $\mathbb{U}_n = \{g^1, g^2, \ldots, g^{\phi(n)-1}, g^{\phi(n)} \equiv 1\}$. This implies that $\mathrm{ord}_n(g) = \phi(n)$. Moreover $\mathbb{U}_n$ is cyclic and g is (one of) its generators. ∎

Example 10.14.1

For $n = 7$ we have $|\mathbb{U}_n| = \phi(7) = 6$, and

$$\mathbb{U}_7 = \{3^1, 3^2, 3^3, 3^4, 3^5, 3^6\} \bmod 7 = \{3,2,6,4,5,1\}.$$

Thus $g = 3$ is a primitive root. The number of primitive roots is $\phi(\phi(n)) = \phi(6) = 2$. The other, second primitive root, is $g = 5$.

$$\mathbb{U}_7 = \{5^1, 5^2, 5^3, 5^4, 5^5, 5^6\} \bmod 7 = \{5,4,6,2,3,1\}.$$

Example 10.14.2

For $n=3$ we have $|\mathbb{U}_n| = \phi(3) = 2$, and

$$\mathbb{U}_3 = \{2^1 \bmod 3, 2^2 \bmod 3\} = \{2,1\}.$$

Thus $g=2$ is a primitive root. The number of primitive roots is $\phi(\phi(3)) = \phi(2) = 1$, and thus there is no other.

Example 10.14.3

For $n=6$ we have $|\mathbb{U}_n| = \phi(6) = 6(1-1/2)/(1-1/3) = 2$, since $6 = 2\cdot 3$ and the number of primitive roots is $\phi(\phi(6)) = \phi(2) = 1$. Moreover

$$\mathbb{U}_6 = \{5,1\} = \{5^1, 5^2\} \bmod 6,$$

we conclude that 5 is the only primitive root mod 6. Note that 3 is the primitive root of $\mathbb{U}_3$. Since $6 = 3\cdot 2$, it might or might not be the primitive root of $\mathbb{U}_6$. If it is not then we try $2+3=5$ and we concluded that indeed the latter was the primitive root.

Example 10.14.4

$\mathbb{U}_4 = \{1,3\}$, and $\mathbb{U}_5 = \{1,2,3,4,\}$,

10.14.6 Jacobi symbol

Definition 10.58 **Jacobi symbol**

Let $n \in \mathbb{Z}$ and $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$. Let n be a product of odd prime numbers, not necessarily distinct,

$$n = p_1 p_1 \dots p_k,$$

then the Jacobi symbol is defined as follows.

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \cdot \left(\frac{a}{p_2}\right) \cdot \dots \cdot \left(\frac{a}{p_k}\right) \cdot .$$

The Jacobi symbol for a prime number n is the Legendre symbol.

Corollary 10.7

Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$. Let n be a product of odd prime numbers, not necessarily distinct,

$$n = p_1 p_1 \dots p_k,$$

and a is a quadratic residue $\bmod n$. Then the following applies.

$$\left(\frac{a}{n}\right) = 1.$$

Proof. If $n = p_1 p_1 \dots p_k$ and given $\gcd(a,n) = 1$, for a a quadratic residue we have $x^2 \equiv a \pmod{n}$ for some x. Since $p_i | n$ and $\gcd(a,n) = 1$ it follows then that $\gcd(a,p_i) = 1$. Moreover $x^2 \equiv a \pmod{p_i}$. Therefore

$$\left(\frac{a}{p_i}\right) = 1, \forall i = 1, \dots, k.$$

Therefore

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \cdot \left(\frac{a}{p_2}\right) \cdot \dots \left(\frac{a}{p_k}\right) = 1 \cdot 1 \cdot \dots \cdot 1 = 1.$$

■

Note that that if $\left(\frac{b}{n}\right) = 1$ it is not necessarily true that b is a quadratic residue $\mod n$.

Theorem 10.10

Let $n, m \in \mathbb{Z}$, $n > 0, m > 0$ be odd (positive) integers. Furthermore let $N, M \in \mathbb{Z}$ such that $\gcd(nm, NM) = 1$. Then the following apply.

$$(a) \left(\frac{N}{n}\right)\left(\frac{N}{m}\right) = \left(\frac{N}{nm}\right),$$

$$(b) \left(\frac{N}{n}\right)\left(\frac{M}{n}\right) = \left(\frac{NM}{n}\right),$$

$$(c) \left(\frac{N^2}{n}\right) = \left(\frac{N}{n^2}\right) = 1,$$

$$(d) \left(\frac{N^2 M}{n^2 m}\right) = \left(\frac{M}{m}\right),$$

and if $N \equiv M \pmod n$, then

$$(e) \left(\frac{N}{n}\right) = \left(\frac{M}{n}\right).$$

Proof.
Left as a problem. ■

10.15 Euler's totient function $\phi(n)$

Exercise 890.

Show in another way that for all $n > 0$

$$n = \sum_{d|n} \phi(d).$$

Solution.

For $d = 1, 2, \ldots, n$ define

$$A_d = \{1 \leq a \leq n : \gcd(a,n) = d\} \tag{10.24}$$

If $d \neq 1$ and $d \nmid n$ then $A_d = \emptyset$. It is straightforward to conclude

$$\sum_{d=1}^{n} |A_d| = n \tag{10.25}$$

From Eq.(10.24) we have the following, given that $d|a$ and thus a/d is an integer and so is n/d.

$$A_d = \{1 \leq a \leq n : \gcd(a,n) = d\} = \{1 \leq \frac{a}{d} \leq \frac{n}{d} : \gcd(\frac{a}{d}, \frac{n}{d}) = 1\}. \tag{10.26}$$

From the latter we conclude $|A_d| = \phi(n/d)$. Eq.(10.26) by way of Eq.(10.25) gives the following.

$$n = \sum_{d=1}^{n} |A_d| = \sum_{d=1}^{n} \phi(n/d) = \sum_{d|n, \frac{n}{d}|n} \phi(n/d) = \sum_{\frac{n}{d}|n} \phi(n/d) = \sum_{D|n} \phi(D) = \sum_{d|n} \phi(d). \tag{10.27}$$

The last few equations involved letter relabelings. ■

Exercise 891.
Find all $n \in \mathbb{N}$ for which $\phi(n)$ is an odd number.

Solution.
$\phi(1) = 1$ and $\phi(2) = 1$. For any other $n > 2$, integer n can be of three forms: (a) $n = 2^k$ for some integer $k > 1$, or (b) $n = p_1^{a_1} p_2^{a_2} \ldots$, where $p_1 = 2$, and p_2 is a odd prime number, or (b) $n = p_1^{a_1} p_2^{a_2} \ldots$, where $p_1 \neq 2$, and p_1, p_2 are all prime numbers. In cases (a), $\phi(n)$ is even. In cases (b), (c) there is a factor $p_2 - 1$ in the former and $p_1 - 1$ or $p_2 - 1$ in the latter of $\phi(n)$. Since the corresponding p_i are odd numbers $p_i - 1$ is even. Therefore $\phi(n)$ is even as well.
The only odd cases left are for $n = 1$ or $n = 2$! ■

Exercise 892.

Show that for any $N,M \in \mathbb{N}$ with $N = p^n$ and $M = p^m$ with p a prime number, and $n,m \in \mathbb{N}$ with $n \geq 0, m \geq 0$, we have that

$$\phi(N \cdot M) = \phi(N) \cdot \phi(M) \cdot \frac{d}{\phi(d)},$$

where $d = \gcd(N,M)$.
Note. $\phi(1) = 1$.

Solution.
If $n = 0$ then $N = p^0 = 1$, $\gcd(N,M) = 1$, and therefore

$$\phi(N \cdot M) = \phi(M) \wedge \phi(N) \cdot \phi(M) \cdot \frac{d}{\phi(d)} = \phi(1)\phi(M) \cdot \frac{1}{\phi(1)} = \phi(M),$$

and the result is true.
If $m = 0$ a similar result is obtained.
If $n > 0, m > 0$ then

$$\phi(N \cdot M) = \phi(p^{n+m}) == p^{n+m-1}(p-1) = \left(p^{k+l-1}(p-1)\right),$$

where $k = \max(n,m)$ and $l = \min(n,m)$. On the other hand

$$\phi(N) \cdot \phi(M) \cdot \frac{d}{\phi(d)} = p^n(p-1)p^m(p-1)\frac{p^l}{p^{l-1}(p-1)} = p^{n+m-1}(p-1).$$

The result is then proven. ■

Exercise 893.
Show that for any $N, M \in \mathbb{N}$ we have that

$$\phi(N \cdot M) = \phi(N) \cdot \phi(M) \cdot \frac{d}{\phi(d)},$$

where $d = \gcd(N, M)$.

Solution.
We use the previous problem. Let $N = p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$ and $M = p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}$ where $n_i \geq 0$ and $m_i \geq 0$. Let $l_i = \min(n_i, m_i)$, $i = 1, \dots, r$. Let $k_i = \max(n_i, m_i)$, $i = 1, \dots, r$. Then $k_i + l_i = n_i + m_i$. Then we have the following.

$$\begin{aligned}
\phi(N) \cdot \phi(M) \cdot \frac{d}{\phi(d)} &= \phi\left(p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}\right) \cdot \phi\left(p_1^{m_1} p_2^{m_2} \dots p_r^{m_r}\right) \cdot \frac{p_1^{l_1} p_2^{l_2} \dots p_r^{l_r}}{\phi(p_1^{l_1} p_2^{l_2} \dots p_r^{l_r})} \\
&= p_1^{n_1-1} p_2^{n_2-1} \dots p_r^{n_r-1} \cdot (p_1-1)(p_2-1)\dots(p_r-1) \\
&\cdot\ p_1^{m_1-1} p_2^{m_2-1} \dots p_r^{m_r-1} \cdot (p_1-1)(p_2-1)\dots(p_r-1) \\
&\cdot\ \frac{p_1^{l_1} p_2^{l_2} \dots p_r^{l_r}}{p_1^{l_1-1}(p_1-1) p_2^{l_2-1}(p_2-1)\dots p_r^{l_r-1}(p_r-1)} \\
&= p_1^{k_1+l_1-2} p_2^{k_2+l_2-2} \dots p_r^{k_r+l_r-2} \cdot (p_1-1)^2(p_2-1)^2\dots(p_r-1)^2 \cdot \frac{p_1 p_2 \dots p_r}{(p_1-1)(p_2-1)\dots(p_r-1)} \\
&= p_1^{k_1+l_1-1} p_2^{k_2+l_2-1} \dots p_r^{k_r+l_r-1} \cdot (p_1-1)(p_2-1)\dots(p_r-1) \\
&= p_1^{n_1+m_1-1} p_2^{n_2+m_2-1} \dots p_r^{n_r+m_r-1} \cdot (p_1-1)(p_2-1)\dots(p_r-1) \\
&= \phi(N \cdot M).
\end{aligned}$$

The result is proven. ■

Exercise 894.

(a) Let $N,m \in \mathbb{N}$ and $N,m > 0$. Show that There are $m \cdot \phi(N)$ integers relatively prime to N among $1,2,\ldots,N \cdot m-1,N \cdot m$.
(b) Use part (a) to conclude that the number of integers relatively prime to N is $\phi(N)$.

Solution.

$$N = p_1^{n_1} p_2^{n_2} \ldots p_{r-1}^{n_{r-1}}, p_r^{n_r},$$

where $n_i > 0$, $r > 0$ and $1 < p_1 < p_2 < \ldots < p_r$. The result uses induction on r.
Base case $r = 0$**.** Then $N = 1$, and $\phi(N) = 1$. True by inspection. The number of integers in $1,2,\ldots,m-1,m$ that are relatively prime to $N = 1$ is m, obviously.
Induction hypothesis: $H(k)$**,** $k < r$. Let us assume that the theorem is true for an

$$n = p_1^{n_1} p_2^{n_2} \ldots p_{r-1}^{n_{r-1}},$$

Then $Nm = (p_r^{n_r} m)n$ and by the induction hypothesis among the integers $1,2,\ldots,N \cdot m-1,N \cdot m$ there are $(p_r^{n_r} m)\,\phi(n)$ integers relatively prime to n such as x, where $\gcd(x,n) = 1$. We now want to find the number of integers y relatively prime to $N = np_r^{n_r}$, where $\gcd(y,N) = 1$. If y is relatively prime to N is also so to n. But if it is to n it might not be so for N: it might be a multiple of p_r. Therefore we need to discard those x that are not relatively prime to $p_r^{n_r}$, i.e. all multiples of p_r. Since p_r is a prime number, the number that are NOT relatively prime to $p_r^{n_r}$ are the multiple of p_r. Among the $(p_r^{n_r} m)\,\phi(n)$ integers $(p_r^{n_r} m)\,\phi(n)/p_r$ of them are NOT relative prime to $p_r^{n_r}$ as they are multiple of p_r and thus the number of integers relatively prime to N in the range $1,2,\ldots,N \cdot m-1,N \cdot m$ is

$$(p_r^{n_r} m)\,\phi(n) - (p_r^{n_r} m)\,\phi(n)/p_r = mp_r^{n_r-1}\,(p_r-1)\,\phi(n) = m\phi(N).$$

The result has been shown.
(b) For the Corollary set $m = 1$ in (a). ■

10.16 Modular equations

Exercise 895.

Modular equations part 1. (Diophantine equation.)

(a) The modular equation

$$ax \equiv b \pmod{n}$$

has a solution if and only if $\gcd(a,n)|b$.

(b) The modular equation

$$ax \equiv 1 \pmod{n}$$

has a solution if and only if $\gcd(a,n) = 1$.

(c) The modular equation, for prime p,

$$ax \equiv 1 \pmod{p}$$

has a solution for x if $p \nmid a$.

Solution.

(a) From $ax \equiv b \pmod{n}$ we have that $ax - b = kn$ i.e. $ax + kn = b$. The solution of this Diophantine equation exists for $\gcd(a,n)|b$. Let $d = \gcd(a,n)$. If x is a solution $n|ax-b$ or $ax-b = kn$ for some integer k. Since $d|a, d|b$ then $d|b$. For the other direction, let $d|b$. Since $d = \gcd(a,n)$ then $1 = \gcd(a/d, n/d)$. Then a/d has an inverse mod n/d and thus $(a/d)m = 1 + (n/d)l$ for integers m, l and thus multiplying by b we have $(a/d)bm = b + (n/d)(lb)$ i.e. $a(bm/d) = b + nlb/d$ i.e. $ax \equiv b \pmod{n}$ with $x = bm/d$. (Note that $d|b$ thus x is an integer.)

(b) The $\gcd(a,n)|b$ of part (a) for $b = 1$ becomes $\gcd(a,n)|1$. Thus $\gcd(a,n) \leq 1$. The gcd is always a positive integer i.e. $\gcd(a,n) \geq 1$. Thus $\gcd(a,n) = 1$ as needed.

(c) By parts (a) and (b) for a solution to exist $\gcd(a,p) = 1$. Since p is a prime its only positive divisors are 1 and p. Given that p cannot divide a, we have that the $\gcd(a,p) = 1$. The result follows. Equivalently, it can be stated as follows. If p is a prime number, the modular equation, for prime p and $p \nmid a$, then there exists an x such that $1 \leq x \leq p-1$ such that the modular equation,

$$ax \equiv 1 \pmod{p}$$

has a solution for x. The x is sometimes denoted as the inverse of a modulo p i.e. $a^{-1} \pmod{p}$. ■

Exercise 896.
(a) Find the multiplicative inverses of 1, 3, 5, 2, 4, 6 mod 11.
(a) Find the multiplicative inverses of 1, 3, 5, 2, 4, 6 mod 9.

Solution.
(a) They are $1,4,9,6,3,2$ respectively.
(b) They are For $1,5,2,4$ they have $1,2,5,7$ as inverses. Integer 3 and 6 have no multiplicative inverses mod 9. ■

10.16.1 Wilson's theorem

Exercise 897.

Wilson's theorem (onlyif). If p is a prime number, then show that

$$(p-1)! \equiv -1 \pmod{p}. \tag{10.28}$$

Solution.

Consider $\{1,2,\ldots p-1\}$. No number x among them is divisible by p thus it is relatively prime to p and thus it has an inverse $\pmod{p}$. It is not possible that x is its own inverse $x^{-1} \pmod{p}$ unless $x=1$ or $x \equiv p-1 \equiv -1 \pmod{p}$ from Theorem 10.33. Thus for the remaining values $x \in \{2,\ldots,p-2\}$, we must have $x \not\equiv x^{-1} \pmod{p}$. Every pair cancels each other i.e. $x \cdot x^{-1} \equiv 1 \pmod{p}$. Thus

$$2 \cdot 3 \cdot 4 \ldots (p-2) \equiv 1 \pmod{p}$$

Restoring the missing 1 and $p-1$ we have

$$1 \cdot 2 \cdot 3 \cdot 4 \ldots (p-2) \cdot (p-1) \equiv p-1 \equiv -1 \pmod{p}$$

■

Exercise 898.

Wilson's theorem (Iff). An integer p is a prime number, if and only if

$$(p-1)! \equiv -1 \pmod{p}. \tag{10.29}$$

Solution.
$\Leftarrow$. If p is not prime then $p = rs$ and $r,s < p$. The term $(p-1)!$ includes all integers $< p$ and thus r and s. This implies that $(p-1)! \equiv 0 \pmod{p}$. There is one exception and that is $p = q^2$. For to have q appearing in the product twice it would mean that q and $2q$ are part of the product, i.e. $2q \leq p-1$. For this to be the case we need $p \geq 4$. Thus we verify by hand exhaustively that for $p = 2,3$ the Theorem is true.
We can also claim that for $p = q^2 > 2q > q$ the $(p-1)! \equiv 0 \pmod{p}$ is still the case. The latter is equivalent to $q^2 > 2q$ i.e. $q > 2$. For $q = 2$ and thus $p = q^2 = 4$ we have $(p-1)! = 1 \cdot 2 \cdot 3 \equiv 2 \pmod{4}$ and thus $(p-1)! \not\equiv -1 \pmod{p}$.
$\Rightarrow$. If p is a prime number, then each $1,2,\ldots,p-1$ has an inverse, and among them, 1 and $p-1$ are their own inverses.

$$x \equiv x^{-1} \pmod{p} \Rightarrow x \cdot x \equiv 1 \pmod{p} \Rightarrow x \cdot x - 1 \equiv 0 \pmod{p} \Rightarrow (x-1)(x+1) \equiv 0 \pmod{p}$$

Thus $p|(x-1)(x+1)$ and since $1 \leq x < p$ we have $p|x-1$ or $p|x+1$ or equivalently $x-1 \equiv 0 \pmod{p}$ or $x+1 \equiv 0 \pmod{p}$ or equivalently $x \equiv 1 \pmod{p}$ or $x \equiv -1 \pmod{p}$.

Thus all the other integers invert in pairs. Thus $(p-1)! \equiv (p-1) \cdot (p-2) \cdot \ldots \cdot 2 \cdot 1 \equiv (p-1) \cdot 1 \equiv -1 \pmod{p}$. ■

Exercise 899.

Let $n > 1$ be an integer and $n \nmid a$. The following are

(a) a is a zero divisor $\pmod{n}$,

(b) a has no inverse $\pmod{n}$,

(c) there exists a $s \in Z$ such that $n \nmid s$ and $as \equiv 0 \pmod{n}$.

Solution.
Statements (a) and (b) are true and equivalent by the prior definition and introduction of unit and zero divisor. Suppose that (b) is true and a has no inverse. By Theorem 10.16 $\gcd(a,n) > 1$. Let $\gcd(a,n) = d > 1$. The $a = dr$ and $n = ds$ for some integer r, s. For $1 << z < n$ we have $z \not\equiv \pmod{n}$. Furthermore, $as = (dr)s = (ds)r = nr \equiv 0 \pmod{n}$. Statement (c) follows from Statement (b).
Suppose that statement (c) is true. There there exists an $s \in Z$ such that $n \nmid a$ and $as \equiv 0 \pmod{n}$. We are going to prove a has no inverse. Let us assume that a has an inverse, then $aa^{-1} \equiv 1 \pmod{n}$, and then

$$0 \equiv as \equiv asa^{-1} \equiv (aa^{-1})s \equiv s \pmod{n}.$$

The latter implies that $n|s$ that contradicts the assumption that $n \nmid s$! Thus a has no inverse and statement (b) is true coming from (c). Thus statements (b) and (c) are equivalent. For prime p, $ab \equiv 0 \pmod{p}$ implies either $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$. Thus for a prime p there no zero divisors other than $0 \pmod{p}$. ■

10.16.2 Fermat's several theorems

Exercise 900.

Fermat's Little theorem. Let $p \in \mathbb{Z}$ be a prime number. Let $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$. Then

$$a^{p-1} \equiv 1 \pmod{p}.$$

A reformulation of this theorem is as follows. Let $f(x) = x^{p-1} - 1$. Then

$$f(x) \equiv 0 \pmod{p},$$

has exactly $p-1$ distinct roots in $(\mathbb{Z}/p\mathbb{Z})^*$. The latter set is $\{1, 2, \ldots, p-1\}$ for a prime p.

Solution.
If p is a prime number, then $\gcd(i,p) = 1 \vee p$. The latter is the case if i is a multiple of p. Say a is not a multiple of p and $\gcd(a,p) = 1$. Consider

$$(1a)_p \cdot (2a)_p \cdot \ldots \cdot ((p-1)a)_p,$$

where $(ia)_p ! = (0)_p$ and thus $ia - 0! = kp$. This is so because otherwise $ia = kp$ would imply that p divides i or p divides a. The former is not possible for prime p since $0 < i < p$, and the latter by way of $\gcd(a,p) = 1$. Furthermore $(ia)_p \neq (ja)_p$ for $i \neq j$ since otherwise $(i-j)a \equiv 0 \pmod{k}p$ and we can use the same consideration as before. Therefore

$$(1a)_p, (2a)_p, \ldots, ((p-1)a)_p$$

is a permutation of $1, 2, \ldots, p-1$ and we obtain the following.

$$(1a)_p \cdot (2a)_p \cdot \ldots \cdot ((p-1)a)_p = ((p-1)! a^{p-1})_p = ((p-1)!)_p,$$

or equivalently,

$$((p-1)! a^{p-1})_p = ((p-1)!)_p \Leftrightarrow ((p-1)! a^{p-1}) \equiv (p-1)! \Leftrightarrow a^{p-1} \equiv 1 \pmod{p}.$$

The last one is obtained given $gcd(p, (p-1)!) = 1$. Completed. ■

Exercise 901.

This has been proven earlier.

Theorem 10.11 **Fermat's Little theorem**

Let $p \in \mathbb{Z}$ be a prime number. For any $a \in \mathbb{Z}$ and $p \nmid a$, we have the following.

$$a^{p-1} \equiv 1 \pmod{p}.$$

In the current formulation $p \nmid a$ is used. In the previous formulation, $p \nmid a$ was replaced by $\gcd(a,p) = 1$.

Solution.
Consider $a, 2a, 3a, \ldots (p-1)a \pmod{p}$. Since $p \nmid a$, all these values are non-zero and distinct mod p. This is because if $ia \equiv ja \pmod{p}$, because $p \nmid a$ it should be $p|i-j$. This means $p \leq |i-j|$. But both i, j are such that $0 \leq |i-j| \leq p-1$ and thus $p \leq p-1$ which is impossible! The $p-1$ values $a, 2a, \ldots (p-1)a \pmod{p}$ can only be the only $p-1$ available $\pmod{p}$ i.e. $1, 2, 3, \ldots, p-1$ (possibly) rearranged. Then taking their product one way or the other,

$$a \cdot 2a \cdot \ldots \cdot (p-1)a \equiv a^{p-1}(p-1)! \equiv 1 \cdot 2 \cdot \ldots \cdot (p-1) \equiv (p-1)! \pmod{p}$$

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}$$

Since p is relatively prime to $1, 2, 3, \ldots, (p-1)$ it is also to $(p-1)!$. Thus $(a^{p-1}-1)(p-1)! \equiv 0 \pmod{p}$. Thus it must be $a^{p-1} - 1 \equiv 0 \pmod{p}$ or equivalently $a^{p-1} \equiv 1 \pmod{p}$ as needed.

For $a \in \mathbb{Z}$, we have $a^p \equiv a \pmod{p}$. Given from Fermat's Little Theorem that $a^{p-1} \equiv a^{p-2}a$ we conclude the following. For $a \in \mathbb{Z}$ such that $p \nmid a$, we have $a^{-1} \equiv a^{p-2} \pmod{p}$. ■

Exercise 902.

Theorem 10.12 **Fermat's theorem**

For any prime number p and any $a \in \mathbb{Z}$ we have

$$a^p \equiv a \pmod{p}. \tag{10.30}$$

If in addition $\gcd(a,p) = 1$, *then*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{10.31}$$

The second part condition $\gcd(a,p) = 1$ of Eq.(10.31) is equivalent to $p \nmid a$ that eventually leads to Fermat's Little theorem of Theorem 10.11.

Solution.
Let $a \geq 0$. We prove the first part by induction on a.
Base case $a = 0$. Then $a^p = 0^p = 0 = a$ and thus $a^p \equiv a \pmod{p}$ by default.
We now assume that for $a \geq 1$ we have

$$(a-1)^p \equiv a-1 \pmod{p}$$

We then have by utilizing the Binomial theorem, and the induction hypothesis above the following.

$$a^p = ((a-1)+1)^p \equiv (a-1)^p + 1^p \equiv (a-1)+1 \equiv a \pmod{p}.$$

Now from $a^p \equiv a \pmod{p}$ we have that there exists a $k \in \mathbb{Z}$ such that

$$a^p - a = kp \Leftrightarrow a(a^{p-1} - 1) = kp,$$

and given $p|kp$ we have $p|a(a^{p-1}-1)$. Since $\gcd(a,p) = 1$, then $p|a^{p-1}-1$ and therefore

$$a^{p-1} \equiv 1 \pmod{p}.$$

■

Exercise 903.

Let $p \in \mathbb{Z}_+$. Show that p is a prime number if and only is $a^{p-1} \equiv 1 \pmod{p}$ for every $a \neq 0$, and $a = 1, \ldots, p-1$. (We can augment a to be any integer $a \in \mathbb{Z}_+$ such that $p \nmid a$.)

Solution.

If p is a prime number this follows from Fermat's little theorem. For the other direction, if the modular equation is true for all a, and p is not a prime number, let q be a divisor of p that is $p = qk$ for some integer k, and $q > 1$. Then $a^{p-1} \equiv 1 \pmod{p}$, implies also $q^{p-1} \equiv 1 \pmod{p}$, and since $q|p$ we also have $q^{p-1} \equiv 1 \pmod{q}$. From the latter, $q|q^{p-1} - 1$ we have $q|1$ or $q \leq 1$ which contradicts $q > 1$. Thus q does not exist and p is a prime number, as needed. ■

Exercise 904.

(a) Show first that if $a,b \in \mathbb{Z}$ such that $a \equiv b \pmod{p^n}$ then $a^p = b^p \pmod{p^{n+1}}$.
(b) Show that for any prime number $p \in \mathbb{N}$ and any $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$ the following applies.

$$a^{(p-1)p^{n-1}} \equiv 1 \pmod{p^n}. \tag{10.32}$$

Solution.
(a) Since $a \equiv b \pmod{p^n}$ we have $a - b = kp^n$ and thus p divides $a-b$ and in fact any p^i, $i \leq n$ divides $a-b$. Thus $a \equiv b \pmod{p}$. Moreover $a^j(a-b) = a^{j+1} - a^j b$ and thus $a^{j+1} \equiv a^j b \pmod{p}$ and also $a^{p-1} \equiv a^j b^{p-1-j} \pmod{p}$. Consider

$$a^p - b^p = (a-b)(a^{p-1} + a^{p-2}b + \ldots + b^{p-1}).$$

Moreover $a^{p-i}b \equiv a^{p-1} \pmod{p}$ and thus

$$(a^{p-1} + a^{p-2}b + \ldots + b^{p-1}) \equiv pa^{p-1} \equiv 0 \pmod{p}.$$

Then from $a \equiv b \pmod{p^n}$ we have that $p^n | a-b$ and $p | (a^{p-1} + a^{p-2}b + \ldots + b^{p-1})$, i.e. $p^{n+1} | a^p - b^p$.
(b) Use induction of $n \geq 1$. For $n = 1$ this follows from the previous problem. Let by the induction hypothesis

$$a^{(p-1)p^{n-2}} \equiv 1 \pmod{p^{n-1}}.$$

Then by part (a) we have

$$a^{((p-1)p^{n-2})p} \equiv 1^p \pmod{p^n}$$

and then

$$a^{(p-1)p^{n-1}} \equiv 1 \pmod{p^n}.$$

■

Exercise 905.
Is 1023 a prime number? Is 511 a prime number?

Solution.
(a) Obviously, 1023 is not a prime number. The sum of its digits is divisible by 3. (See a prior problem.)
(b) 511 is not a prime number. We show that if 511 was a prime number by Fermat's Little Theorem, for every a such that $\gcd(a,p)=1$ we would have have $a^{p-1}\equiv 1 \pmod{p}$ or equivalently $a^{510}\equiv 1 \pmod{511}$.
We observe

$$2^9 = 512 \equiv 1 \pmod{511}.$$

Furthermore $2^6 \equiv 64 \pmod{511}$. Moreover $510 = 56\cdot 9+6$. Therefore

$$2^{510} = 2^{56\cdot 9+6} = (2^9)^{56}\cdot 2^6 \equiv 1^{56}\cdot 64 \equiv 64 \pmod{511}.$$

Give that $\gcd(2,511)=1$ this mean 511 is not a prime number. ■

Exercise 906.
Is 2047 a prime number? Explain.

Solution.
If we pick $a = 2$ we have

$$a^{11} = 2^{11} = 2048 \equiv 1 \pmod{2047},$$

and

$$a^{2046} = 2^{2046} = 2^{186 \cdot 11} \equiv 1 \pmod{2047},$$

a misleading indicator. But if we pick $a = 3$ we have the following.

$$\begin{aligned}
3^0 &\equiv 1 \pmod{2047} \\
3^1 &\equiv 3 \pmod{2047} \\
3^2 &\equiv 9 \pmod{2047} \\
3^4 &\equiv 81 \pmod{2047} \\
3^8 &\equiv 420 \pmod{2047} \\
3^{10} &\equiv 1733 \pmod{2047} \\
3^{16} &\equiv 176400 \equiv 358 \pmod{2047} \\
3^{32} &\equiv 1250 \pmod{2047} \\
3^{64} &\equiv 639 \pmod{2047} \\
3^{128} &\equiv 968 \pmod{2047} \\
3^{256} &\equiv 1545 \pmod{2047} \\
3^{512} &\equiv 223 \pmod{2047} \\
3^{1024} &\equiv 601 \pmod{2047}.
\end{aligned}$$

We then note the following.

$$\begin{aligned}
3^{2046} &= 3^{2016} \cdot 3^{30} = 3^{1024+512+256+128+64+32} \cdot 3^{30} \\
&\equiv 601 \cdot 223 \cdot 1545 \cdot 968 \cdot 639 \cdot 1250 \cdot 3^{30} \pmod{2047} \\
&\equiv 1545 \cdot 3^{30} \pmod{2047} \\
&\equiv 1545 \cdot 3^{16} \cdot 3^{10} \cdot 3^4 \pmod{2047} \\
&\equiv 1545 \cdot 358 \cdot 1733 \cdot 81 \pmod{2047} \\
&\equiv 1013 \pmod{2047} \\
&\not\equiv 1 \pmod{2047}.
\end{aligned}$$

One can further confirm the following.

$$\begin{aligned}
3^{2047} &= 3^{2016} \cdot 3^{31} \\
&= 3^{1024+512+256+128+64+32} \cdot 3^{30} \\
&\equiv 601 \cdot 223 \cdot 1545 \cdot 968 \cdot 639 \cdot 1250 \cdot 3^{30} \pmod{2047} \\
&\equiv 1545 \cdot 3^{31} \pmod{2047} \\
&\equiv 1545 \cdot 3^{16} \cdot 3^8 \cdot 3^4 \cdot 3^2 \cdot 3^1 \pmod{2047} \\
&\equiv 1545 \cdot 358 \cdot 420 \cdot 81 \cdot 9 \cdot 3 \pmod{2047} \\
&\equiv 1545 \cdot 1099 \pmod{2047} \\
&\equiv 992 \pmod{2047} \\
&\not\equiv 3 \pmod{2047}.
\end{aligned}$$

■

Exercise 907.

(a) Consider the integers $\pmod 8$. We have $3^2 \equiv 1 \pmod 8$. We also have $5^2 \equiv 1 \pmod 8$. These are two square roots of $1 \pmod 8$.

(b) Can you find others (e.g. 1, 7)?

(c) Moreover $4 \equiv -4 \pmod 8$. $a \equiv -a \pmod n$ is equivalent to $2a \equiv 0 \pmod n$. If $n = 2a$ this is trivially true. If n is odd, then $\gcd(2,n) = 1$ and thus $n|a$. In the latter case $a \equiv 0 \pmod n$.

(d) Show the following. If p is an odd prime ($p \neq 2$) and $p \nmid a$ then the equation

$$x^2 \equiv a \pmod p \tag{10.33}$$

has either exactly two distinct roots or no roots at all.

Solution.
If there are NO roots to the modular equation the proof is complete.
Otherwise let z be a solution i.e. $z^2 \equiv a \pmod p$. Since $-z$ is such that $(-z)^2 \equiv z^2 \equiv a \pmod p$, then $-z$ is also a solution. Is it $z \equiv -z \pmod p$? This is so if p is an even number as it was shown prior to the statement of Eq. (10.33). It is also possible that p is odd but then it must divide a. However the preconditions of the theorem disallow the former (an even number cannot be the case as p is given to be an odd prime) and the latter (being an even number and $p|a$ is not possible, since p is an odd prime number and $p \nmid a$).
Therefore there are two distinct solutions $z, -z \pmod p$ if one of them (say z) exists. Does there exist a third (or fourth etc) solution? Let us call it w. Then $w^2 \equiv z^2 \equiv a \pmod p$. This implies $w^2 - z^2 \equiv 0 \pmod p$. Then $(w-z)(w+z) \equiv 0 \pmod p$. Then $p|w-z$ or $p|w+z$. In other words $w \equiv z$ or $w \equiv -z$. There are two and only two solutions then. No third or more! ■

Exercise 908.
Show that for $a,b \in \mathbb{U}_n$ we have that $ab \in \mathbb{U}_n$ and also $a^{-1}, b^{-1}, (ab)^{-1} \in \mathbb{U}_n$.

Solution.
Starting with the last result if $a \in \mathbb{U}_n$ it means that $aa^{-1} \equiv 1 \pmod{n}$. Moreover $a^{-1}a^{-1^{-1}} \equiv 1 \pmod{n}$. Thus $a^{-1} \in \mathbb{U}_n$. Similarly, same for b. For a,b let their inverses be a^{-1}, b^{-1} respectively. From the already established result $a^{-1}, b^{-1} \in \mathbb{U}_n$. Consider (ab). Since

$$(ab)(b^{-1}a^{-1})a(bb^{-1})a^{-1} \equiv a \cdot 1 \cdot a^{-1} \equiv 1 \pmod{n}$$

it shows that $ab \in \mathbb{U}_n$. The result for $(ab)^{-1}$ can be proven likewise. ■

10.16.3 RSA

Exercise 909.

RSA. Let $N = pq$, $ed \equiv 1 \pmod{(p-1)(q-1)}$, where $0 < d < (p-1)(q-1)$ and (d,N) is the decryption (secret) key, (e,N) is the encryption (public) key. Proof that it works.

Solution.
Let p,q be distinct primes $N = pq$, $\gcd(a,N) = 1$. Then $a^{(p-1)(q-1)} \equiv 1 \pmod{N}$. For a message $0 < M < n$, encryption is

$$S = M^e \bmod n,$$

and decryption is

$$S^d \equiv M^{ed} \equiv M^{1+k(p-1)(q-1)} \equiv M \cdot M^{k(p-1)(q-1)} \pmod{n} \equiv M \cdot 1 \pmod{n} \equiv M$$

since $\gcd(M,n) = 1$. one needs to check $M < n$ or that $\gcd(M,n) = 1$, or change modulus N. ■

10.17 Euler's theorem

Exercise 910.

Theorem 10.13 **Euler's theorem**

For $a \in Z$, $a > 1$, if $\gcd(a,n) = 1$ then $a^{\phi(n)} \equiv 1 \pmod{n}$.

Solution.
First proof. Let

$$a_1, \ldots, a_{\phi(n)} \pmod{n}$$

be the list of units. Multiply each one of the elements with a, where $\gcd(a,n) = 1$. Since it is so $ax \equiv 1 \pmod{n}$. Moreover a_i are units thus $a_i x_i \equiv 1 \pmod{n}$ as well. The aa_i are such that $(aa_i)(xx_i) \equiv (ax)(a_i x_i) \equiv 1 \pmod{n}$. That is all aa_i are units.

$$aa_1, aa_2, \ldots, aa_{\phi(n)} \pmod{n}.$$

Moreover $aa_i \not\equiv aa_j \pmod{n}$. This is because if $aa_i \equiv aa_j \pmod{n}$ this would imply $xaa_i \equiv xaa_j \pmod{n}$ i.e. $a_i \equiv a_j \pmod{n}$. If all of them are in the integer interval $[1, n-1]$, then this implies $a_i = a_j$, i.e. $i = j$. That is the two lists above are the same up to a reordering of the same elements. Thus

$$aa_1 aa_2, \ldots, aa_{\phi(n)} \equiv a_1 \ldots a_{\phi(n)} \pmod{n}$$

This means

$$a^{\phi(n)} a_1 a_2 \ldots a_{\phi(n)} \equiv a_1 a_2 \ldots a_{\phi(n)} \pmod{n}$$

Using cancellation (or multiplication with $a_1^{-1} a_2^{-1} \ldots$) we once more conclude that $a^{\phi(n)} \equiv 1 \pmod{n}$. (Alternatively $\gcd(a_i, n) = 1$ and $\gcd(\prod_i a_i, n) = 1$. By Bezout's result there exists an x such that $x \prod_i a_i \equiv 1 \pmod{n}$. We then multiply the left and right-hand side about with x.)
Second proof. A second slightly easier proof can be obtained by way of Eq.(10.32).
Let $n = p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$ where $n_i \geq 0$ and $r > 0$. By way of Eq.(10.32) we have the following for all $i = 1, \ldots, r$, for an $a \in \mathbb{N}$ and $\gcd(a,n) = 1$.

$$a^{p_i^{n-1}(p_i - 1)} \equiv 1 \pmod{p_i^{n_i}}. \tag{10.34}$$

Raising equation Eq.(10.34) to the power $\phi(n/p_i^{n_i})$ we obtain the following.

$$\left(a^{p_i^{n-1}(p_i-1)}\right)^{\phi(n/p_i^{n_i})} \equiv 1 \pmod{p_i^{n_i}} \Leftrightarrow a^{\phi(n)} \equiv 1 \pmod{p_i^{n_i}}. \tag{10.35}$$

Given that p_i are relative prime to each other the last one implies the following, for $n = p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}$.

$$a^{\phi(n)} \equiv 1 \pmod{n}. \tag{10.36}$$

■

Exercise 911.
Consider $n > 1$, with n odd. Determine the remainder of the division by n of $N = 2^{\phi(n)-1}$.

Solution.
We use Euler's theorem. Pick $a = 2$. Since n is odd $\gcd(a,n) = \gcd(2,n) = 1$. Then by Euler's theorem

$$2^{\phi(n)} \equiv 1 \pmod{n} \Leftrightarrow 2^{\phi(n)} \equiv 1+n \pmod{n} 2^{\phi(n)-1} \equiv \frac{1+n}{2} \pmod{n}$$

Note that n being odd $n+1$ is even thus $(n+1)/2$ is an integre. Given that $(n+1)/2 < n$ we conclude that the desired remainder is $(n+1)/2$. ■

Exercise 912.
If $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$, then $\phi(n) = n(1-1/p_1)\dots(1-1/p_k)$.

Solution.
By way of Eq. (10.38), $\phi(n) = \phi(p_1^{a_1})\dots\phi(p_k^{a_k})$. Furthermore, from Eq. (10.46) we have

$$\phi(n) = \phi(p_1^{a_1})\dots\phi(p_k^{a_k}) = p_1^{a_1} - p_1^{a_1-1} \dots p_k^{a_k} - p_k^{a_k-1} = n(1-1/p_1)\dots(1-1/p_k).$$

■

10.18 Modular equations continued

Exercise 913.

Modular equations part 2.

The modular equation

$$ax \equiv b \pmod{n}$$

has a solution if and only if $\gcd(a,n)|b$. Let $\gcd(a,n) = d$, and $a_1 = a/d, b_1 = b/d, n_1 = n/d$. If x is a solution of the modular equation then, we have the following

$$x \equiv b_1 a_1^{\phi(n_1)-1} \pmod{n_1}.$$

Solution.

From $ax \equiv b \pmod{n}$ we have that $ax - b = kn$ i.e. $ax + kn = b$, for some $k \in \mathbb{Z}$. The solution of this Diophantine equation exists for $\gcd(a,n)|b$. Let $\gcd(a,n) = d$, and $a_1 = a/d, b_1 = b/d, n_1 = n/d$.
If $d|b$ then we have the following for some $k \in \mathbb{Z}$.

$$\begin{aligned} ax &\equiv b \pmod{n} \Leftrightarrow \\ (a_1 d)x &\equiv b_1 d \pmod{n_1 d} \Leftrightarrow \\ (a_1 d)x - b_1 d &= n_1 dk \Leftrightarrow \\ a_1 x - b_1 &= n_1 k \Leftrightarrow \\ a_1 x &\equiv b_1 \pmod{n_1}. \end{aligned}$$

By Euler's theorem it is $a_1^{\phi(n_1)} \equiv 1 \pmod{n_1}$. We multiply the modular equation above by $a_1^{\phi(n_1)-1}$.

$$\begin{aligned} a_1 x &\equiv b_1 \pmod{n_1} \\ a_1 x a_1^{\phi(n_1)-1} &\equiv b_1 a_1^{\phi(n_1)-1} \pmod{n_1} \\ x a_1^{\phi(n_1)} &\equiv b_1 a_1^{\phi(n_1)-1} \pmod{n_1} \\ x &\equiv b_1 a_1^{\phi(n_1)-1} \pmod{n_1}. \end{aligned}$$

The result follows. ■

10.18.1 Chinese remainder theorem

Exercise 914.

System of two equations: Chinese remainder theorem. Let $a_1, a_2 \in Z$. Let $n_1, n_2 \in \mathbb{N}$ with $n_1, n_2 > 1$, Let d be the g.c.d of n_1, n_2 such that $d = \gcd(n_1, n_2)$. If $d = 1$ and thus n_1, n_2 are relatively prime, then the system of two equations

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \end{aligned} \tag{10.37}$$

has a unique solution $x \bmod (n_1 \cdot n_2)$. Moreover in $\mathbb{Z}$ any two solutions x, y are such that $x \equiv y \pmod{n_1 n_2}$.

Solution.
Let $1 = \gcd(n_1, n_2)$. Then there exist m_1 such that $m_1 n_1 \equiv 1 \pmod{n_2}$, and m_2 such that $m_2 n_2 \equiv 1 \pmod{n_1}$. This follows from the fact that $\gcd(n_1, n_2) = 1$, and thefore

$$m_1 n_1 - 1 = n_2 k_2, \quad m_2 n_2 - 1 = n_1 k_1$$

have solution for n_1, n_2 by the extended GCD algorithm. Consider then integer X as follows

$$X = (a_1 m_1 n_1 + a_2 m_2 n_2) \bmod n_1 n_2.$$

It is obvious that

$$X \pmod{n_1} \equiv 0 + a_2 m_2 n_2 \equiv a_2 \pmod{n_1},$$

since $m_2 n_2 \equiv 1 \pmod{n)_1}$. Similarly,

$$X \equiv a_1 \pmod{n_2}.$$

It then suffices to show that there is no $Y \bmod (n_1 \cdot n_2)$ For if there is a Y such that

$$Y \equiv a_2 \pmod{n_1}, \quad Y \equiv a_1 \pmod{n_2},$$

then $n_1 | Y - X$ and $n_2 | Y - X$. Since $\gcd(n_1, n_2) = 1$, then $n_1 n_2 | Y - X$ which implies $n_1 n_2 \leq Y - X$ or $X = Y$. But $Y - X$ or $X - Y$ is between 0 and $n_1 n_2 - 1$ contradicting the conclusion $n_1 n_2 \leq Y - X$. The only alternative left $X = Y$ or $X \equiv Y \pmod{n_1 n_2}$. ■

Exercise 915.
If $\gcd(m,n)=1$ then

$$\phi(mn)=\phi(m)\phi(n). \tag{10.38}$$

Solution.
If a is a unit $\pmod{mn}$ it means $\gcd(a,mn)=1$. Using a Chinese remainder theorem-based method consider function g defined on $\mathbb{Z}_{mn} \Rightarrow \mathbb{Z}_m \times \mathbb{Z}_n$.

$$g(A)=(A \pmod m,\ A \pmod n).$$

we will show that g is a one-to-one and onto bijection. Note that if A is a unit $\pmod{mn}$ then by Proposition 10.47 it is a unit $\pmod m$ and a unit $\pmod n$. Consider $G(A_1)=G(A_2)$. Then $A_1 \equiv A_2 \pmod m$ and $A_1 \equiv A_2 \pmod n$. Thus $m|A_1-A_2$ and $n|A_1-A_2$. If $\gcd(m,n)=1$ as it is, then $mn|A_1-A_2$ as well implying $A_1 \equiv A_2 \pmod{mn}$. Say A_1 is a unit $\pmod m$ and A_2 a unit $\pmod n$. Then from $\gcd(m,n)=1$ and Eq.(10.37) there is a unique $A \pmod{mn}$ such that $A \equiv A_1 \pmod m$ and $A \equiv A_2 \pmod n$. Thys $g(A)=(A_1,A_2)=(A \pmod m, A \pmod n)$. Thus the two sets $\mathbb{Z}_{mn}$ and $\mathbb{Z}_m \times \mathbb{Z}_n$ have the same number of elements thus $\phi(mn)=\phi(m)\phi(n)$. ■

10.18.2 System of two equations

Exercise 916.

System of two equations. Let $a_1, a_2 \in Z$. Let $n_1, n_2 \in \mathbb{N}$ with $n_1, n_2 > 1$, Let d be the g.c.d of n_1, n_2 such that $d = \gcd(n_1, n_2)$. If $d|a_1 - a_2$ or equivalently $a_1 \equiv a_2 \pmod{d}$, then the system of two equations

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \end{aligned} \tag{10.39}$$

has a unique solution $x \pmod{\operatorname{lcm}(n_1, n_2)}$. If $a_1 \not\equiv a_2 \pmod{d}$, or equivalently $d \nmid a_1 - a_2$, then the system of equations has no solution in x.

Note. This case deals with a $d > 1$; in the Chinese Remainder Theorem we assume $d = 1$.

Solution.

Let $d = \gcd(n_1, n_2)$. Then $n_1 = dk_1$ and $n_2 = dk_2$ for some $k_1, k_2 \in \mathbb{Z}$. Let the system of Eq.(10.39) has one solution x. Then $x \equiv a_1 \pmod{n)_1}$ and $x \equiv a_2 \pmod{n)_2}$. From the two we have for some $l_1, l_2 \in \mathbb{Z}$. $x - a_1 = n_1 l_1$ and $x - a_2 = n_2 l_2$. From the former we obtain

$$x - a_1 = n_1 l_1 = (dk_1) l_1 = d(k_1 l_1),$$

thus concluding $d|x - a_1$. Similarly we obtain $d|x - a_1$. Then $d|a_1 - a_2$ and therefore

$$a_1 \equiv a_2 \pmod{d}.$$

Furthermore if $a_1 \not\equiv a_2 \pmod{d}$, then the system of Eq.(10.39) has no solution in x, since otherwise $a_1 \equiv a_2 \pmod{d}$.

Assume in the remainder that $a_1 \equiv a_2 \pmod{d}$ and thus the system of Eq.(10.39) has one solution x. Using the tools of the previous proof

$$x - a_1 = n_1 l_1 \wedge x - a_2 = n_2 l_2$$

and by subtracting one from the other we obtain the following.

$$a_1 - a_2 = n_2 l_2 - n_1 l_1 \Leftrightarrow n_1 l_1 = a_2 - a_1 + n_2 l_2 \Leftrightarrow n_1 l_1 \equiv a_2 - a_1 \pmod{n_2}$$

The latter modular equation has a solution l_1 for $\gcd(n_1, n_2) = d$ is such that $d|a_2 - a_1$ by way of $a_1 \equiv a_2 \pmod{d}$. From the previous problem the solution l_1 is as follows.

$$l_1 \equiv \frac{a_2 - a_1}{d} \left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})-1} \pmod{\frac{n_2}{d}}.$$

We then have the following also using that $x - a_1 = n_1 l_1$.

$$\begin{aligned} l_1 &\equiv \frac{a_2 - a_1}{d} \left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})-1} \pmod{\frac{n_2}{d}} \Leftrightarrow \\ l_1 &= \frac{a_2 - a_1}{d} \left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})-1} + m_2 \frac{n_2}{d} \Leftrightarrow \\ x &= a_1 + n_1 \cdot \frac{a_2 - a_1}{d} \left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})-1} + n_1 \cdot m_2 \frac{n_2}{d} \Leftrightarrow \\ x &= a_1 + (a_2 - a_1) \left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})} + m_2 \frac{n_1 \cdot n_2}{d} \Leftrightarrow \\ x &= A + m_2 \frac{n_1 \cdot n_2}{d} \Leftrightarrow \\ x &= A + m_2 \operatorname{lcm}(n_1, n_2) \Leftrightarrow \\ x &\equiv A \pmod{\operatorname{lcm}(n_1, n_2)}, \end{aligned}$$

where A is given by the following expression.

$$A = a_1 + (a_2 - a_1)\left(\frac{n_1}{d}\right)^{\phi(\frac{n_2}{d})}.$$

■

10.18.3 Generalized chinese remainder theorem

Exercise 917.

Modular system of equations: Chinese remainder theorem of more than 2 equations. Let $n_1,\ldots,n_k$ be pairwise prime numbers i.e. $\gcd(n_i,n_j)=1$ for $i\neq j$, $i,j=1,\ldots,k$. Let $a_1,\ldots,a_k\in\mathbb{Z}$. There is a unique A mod $n_1\ldots n_k$ that it satisfies all of the modular equations below.

$$\begin{aligned} A &\equiv a_1 \pmod{n_1} \\ A &\equiv a_2 \pmod{n_2} \\ &\ldots \\ A &\equiv a_k \pmod{n_k} \end{aligned} \tag{10.40}$$

Moreover if there are two solutions A,a, then

$$A\equiv a \pmod{n_1\ldots n_k} \Rightarrow A\equiv a \pmod{N},$$

where $N=n_1n_2\ldots n_k$.

Solution.
Let N_j, for $j=1,\ldots,k$ contain all n_i except n_j. That is

$$N_j=n_1\ldots n_{j-1}n_{j+1}\ldots n_k.$$

We have $n_i|N_j$ for all $i\neq j$. We have that $\gcd(N_j,n_j)=1$. This is because $\gcd(n_j,n_i)=1$ for all $i\neq j$ as they are pairwise prime . Since $\gcd(N_j,n_j)=1$ we have that there exists an integer x_j such that

$$N_jx_j\equiv 1 \pmod{n_j},$$

for all $j=1,\ldots,k$. We then form

$$A=a_1N_1x_1+\ldots+a_iN_ix_i+\ldots+a_kN_kx_k.$$

Consider n_j and N_j. It is $n_i|N_j$ for all $i\neq j$. Thus all the terms $a_jN_jx_j$ are multiples of n_i for $j\neq i$. For the term $a_iN_ix_i$ this is not the case as $\gcd(n_i,N_i)=1$. But we have that $N_ix_i\equiv 1 \pmod{n_i}$. Thus $a_iN_ix_i\equiv a_i \pmod{n_i}$. The second equivalence below follows

$$\begin{aligned} A &\equiv a_1\cdot 0+a_2\cdot 0+\ldots+a_i\cdot N_i\cdot x_i+\ldots+a_k\cdot 0 \pmod{n_i} \\ &\equiv a_i \pmod{n_i} \end{aligned}$$

This is true for all i and the following applies.

$$\forall 1\leq i\leq k: \quad A\equiv a_i \pmod{n_i}.$$

Say that there is another solution a i.e. $a\equiv A\equiv a_i \pmod{n_i}$ for all i. This would mean that $n_i|A-a$. Since n_i are pairwise prime then by way of "if $\gcd(a,b)=1$ and $a|c$ and $b|c$ then $ab|c$", we have

$$n_1n_2\ldots n_k|A-a.$$

This means $A-a\equiv 0 \pmod{n_1n_2\ldots n_k}$. ■

Exercise 918.
Let $n_1,\ldots,n_k$ be pairwise prime numbers i.e. $\gcd(n_i,n_j)=1$ for $i\neq j,\ i,j=1,\ldots,k$. Then $A\equiv a\pmod{n_i}$, $i=j=1,\ldots,k$ if and only if

$$A\equiv a \pmod{N},$$

where $N=n_1n_2\ldots n_k$.

Solution.
The proof follows directly from the previous problem. ■

Exercise 919.
Let $n_1, \ldots, n_k$ be pairwise prime numbers i.e. $\gcd(n_i, n_j) = 1$ for $i \neq j$, $i, j = 1, \ldots, k$. Let $a_1, \ldots, a_k \in \mathbb{Z}$, and let us assume that there is a solution for the system of modular equations below.

$$\begin{aligned} A &\equiv a_1 \pmod{n_1} \\ A &\equiv a_2 \pmod{n_2} \\ &\ldots \\ A &\equiv a_k \pmod{n_k} \end{aligned} \tag{10.41}$$

Furthermore, let $N = n_1 n_2 \ldots n_k$. Show that the solution is as follows.

$$A = \sum_{i=1}^{k} a_i N_i x_i \pmod{N}$$

where $N_i = N/n_i$ and $N_i x_i \equiv 1 \pmod{n_i}$, $i = 1, 2, \ldots, k$.

Solution.
It follows directly from the proof of existence of Eq.(10.40). ■

Exercise 920.
Find all integers A such that

$$\begin{aligned} A &\equiv 2 \pmod 3 \\ A &\equiv 3 \pmod 5 \\ A &\equiv 2 \pmod 7. \end{aligned}$$

This is the original problem proposed by Sunzi that appeared in the book Sunzi Suanjing. A solution though was not proposed there.

Solution.
Let $N = 3\cdot 5\cdot 7 = 105$. Let $n_1 = 3, n_2 = 5, n_7 = 7$. Let $N_1 = N/n_1 = 35$, and similarly $N_2 = 21$ and $N_3 = 15$. Moreover $N_1 \equiv 2 \pmod{n_1}$, $N_2 \equiv 1 \pmod{n_2}$, $N_3 \equiv 1 \pmod{n_3}$. Then $x_i N_i \equiv 1 \pmod{n_i}$ yields $x_1 = 2, x_2 = 1, x_3 = 1$. Then

$$A = 2\cdot 35\cdot 2 + 3\cdot 21\cdot 1 + 2\cdot 15\cdot 1 \pmod N = 233 \pmod{105} = 23.$$

■

Exercise 921.
Find all integers A such that

$$\begin{aligned} A &\equiv 1 \pmod 2 \\ A &\equiv 2 \pmod 3 \\ A &\equiv 3 \pmod 5 \end{aligned}$$

Solution.
Let $n_1 = 2, n_2 = 3, n_3 = 5$.
Then $N_1 = 3 \cdot 5 = 15$, $N_2 = 2 \cdot 5 = 10$, $N_3 = 2 \cdot 3 = 6$.

$$\begin{aligned} N_1 x_1 &\equiv 1 \pmod n_1 \Rightarrow 15x_1 \equiv 1 \pmod 2 \Rightarrow x_1 = 1 \\ N_2 x_3 &\equiv 1 \pmod n_2 \Rightarrow 10x_2 \equiv 1 \pmod 3 \Rightarrow x_2 = 1 \\ N_3 x_4 &\equiv 1 \pmod n_3 \Rightarrow 6x_3 \equiv 1 \pmod 5 \Rightarrow x_3 = 1 \end{aligned}$$

Then $n_1 n_2 n_3 = 2 \cdot 3 \cdot 5 = 30$,

$$A = a_1 N_1 x_1 + a_2 N_2 x_2 + a_3 N_3 x_3 = 1 \cdot 15 \cdot 1 + 2 \cdot 10 \cdot 1 + 3 \cdot 6 \cdot 1 = 15 + 20 + 18 \pmod{30} = 23$$

Thus one solution is 23. Another one is $23 + 30$, another one is $23 + 60$, and so on. ■

10.18.4 Modular system of polynomial equations

Exercise 922.

Modular system of polynomial equations. Let $f(x)$ be a polynomial of degree n of integer coefficients. A polynomial congruence equation is $f(x) \equiv 0 \pmod{m}$ for some integer m. Let $n_1, \ldots, n_k$ be pairwise prime numbers i.e. $\gcd(n_i, n_j) = 1$ for $i \neq j$, $i, j = 1, \ldots, k$. Let $N = n_1 n_2 \ldots n_k$. Then integer a is a solution of

$$f(x) \equiv 0 \pmod{N},$$

if and only if a is a solution of the system of the following polynomial equations.

$$\begin{aligned} f(a) &\equiv 0 \pmod{n_1} \\ f(a) &\equiv 0 \pmod{n_2} \\ &\ldots \\ f(a) &\equiv 0 \pmod{n_k} \end{aligned} \tag{10.42}$$

Moreover if there are two solutions a, b, then

$$a \equiv b \pmod{n_1 \ldots n_k} \Rightarrow a \equiv b \pmod{N},$$

Solution.
$\Rightarrow$. If a is a solution to the congruence equation

$$f(x) \equiv 0 \pmod{N},$$

then $f(a) \equiv 0 \pmod{N}$. Since $n_i | N$ for all $i = 1, \ldots, k$, the $f(a) \equiv 0 \pmod{n_i}$, and ther result follows.
$\Leftarrow$. If a is a solution of the sytem of equation (10.42), then set $A = f(a)$

$$\begin{aligned} A = f(a) &\equiv 0 \pmod{n_1} \\ A = f(a) &\equiv 0 \pmod{n_2} \\ &\ldots \\ A = f(a) &\equiv 0 \pmod{n_k}, \end{aligned} \tag{10.43}$$

and by the way of Eq.(10.40) that is the Chinese remainder theorem we obtain that there is an $A = f(a)$, such that

$$A = f(a) \equiv 0 \pmod{N},$$

as needed. ■

Exercise 923.
Show that the modular equation $x^2 \equiv a \pmod{p}$, where p is a prime number has either 0 solutions or two (congruence) solutions. (Integer a is this $0 < a < p$.)

Solution.
If the equation has zero solutions, we are done. Otherwise let it have one and let it be x. Then we have the following

$$(p-x)^2 \equiv p^2 - 2px + x^2 \equiv x^2 \equiv a \pmod{p},$$

and thus a second solution $p - x \neq x$ has been found. We show that there are no more than them. Let x, y be two solutions. We have the following.

$$x^2 \equiv a \pmod{p}, \quad y^2 \equiv a \pmod{p}$$

which implies

$$x^2 - y^2 \equiv 0 \pmod{p},$$

or equivalently $p|x^2 - y^2$ or $p|(x-y)(x+y)$. For prime p this means $p|x-y$ or $p|x+y$ in other words $x \equiv \pm y \pmod{p}$. Our solution $p - x$ is such that

$$p - x \equiv -x \pmod{p},$$

and is the complementary solution of x. ■

Exercise 924.

For an a such that

$$x^2 \equiv a \pmod{p}$$

has a solution (or as we have shown two then), we call a a quadratic residue. Otherwise, if it has no solution then a is a quadratic non-residue. (We use qr or qnr or q.r. and q.n.r. respectively.)
(a) Find all the quadratic residues and quadratic non-residues mod 11.
(b) For prime p show that there $(p-1)/2$ q.r. and $(p-1)/2$ q.n.r.

Solution.
(a) We square by hand all integers from 1 to 10 mod 11. There are five pairs.

$$\begin{aligned}
1^2 &\equiv 1 \pmod{11} \quad , \quad 10^2 \equiv 1 \pmod{11} \\
2^2 &\equiv 4 \pmod{11} \quad , \quad 9^2 \equiv 4 \pmod{11} \\
3^2 &\equiv 9 \pmod{11} \quad , \quad 8^2 \equiv 9 \pmod{11} \\
4^2 &\equiv 5 \pmod{11} \quad , \quad 7^2 \equiv 5 \pmod{11} \\
5^2 &\equiv 3 \pmod{11} \quad , \quad 6^2 \equiv 3 \pmod{11}
\end{aligned}$$

We thus concludes $1,3,4,5,9$ are the quadratic residues. The $2,6,7,8$ are the quadratic non-residues.
(b) We have $p-1$ equations $x^2 \equiv a \pmod{p}$, for $p-1$ values of a. Let q of them have 2 solutions and $p-1-q$ of them have 0 solutions. Let the equations with two solutions be $x^2 \equiv a_i \pmod{p}$, for $i = 1,\ldots,q$. The there are $2q$ values x that satisfy those equation for the $1,\ldots,p-1$ values x. Each such x satisfy one such equation for some a_i. Therefore $2q = p-1$ and therefore $q = (p-1)/2$. If q are the q.r. mod p, then $p-1-q = (p-1)/2$ are the number of q.n.r. ■

10.19 Diophantine equations continued

Exercise 925.

Diophantine equation for $n > 2$. Let $n \geq 2$. Show that the diophantine equation

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b \tag{10.44}$$

has an integer solution $(x_1, x_2, \ldots, x_n) \in \mathbb{Z} \times \ldots \times \mathbb{Z} = \mathbb{Z}^n$ provided that $(a_1, a_2, \ldots, a_n) = 1$ that is, the gcd of $a_1, a_2, \ldots, a_n$ is equal to 1. Moreover all ineger solutions can be expressed in terms of $n-2$ integer parameters.

Solution.
The proof is by induction on n. The $n = 1$ case is trivially true as $(a_1) = 1$ implies $a_1 = 1$ or $a_1 = -1$ and thus the integer solution for x_1 is $x_1 = b$ or $x_1 = -b$.
For $n \geq 2$ assume that it is true for $n-1$ and let $H(n-1)$ is the hypothesis and in the inductive step show that $H(n)$ is true. Hypothesis (assumption) $H(n-1)$ is stated as follows.

$$H(n-1) : a_1x_1 + a_2x_2 + \ldots + a_{n-1}x_{n-1} = b$$

has an integer solution $(x_1, x_2, \ldots, x_{n-1}) \in (\mathbb{Z})^{n-1}$ provided that $(a_1, a_2, \ldots, a_{n-1}) = 1$, and expressed in terms of $n-2$ integer parameters. We are going to show $H(n)$ that is, Eq.(10.44) for n variables. Let

$$a_1x_1 + a_2x_2 + \ldots + a_{n-1}x_{n-1} + a_nx_n = b,$$

and by taking $\pmod d$ of both sides, we obtain the following.

$$a_1x_1 + a_2x_2 + \ldots + a_{n-1}x_{n-1} + a_nx_n \equiv b \pmod d.$$

By $H(n-1)$ we have that d divides all $a_1, \ldots a_{n-1}$ as their common divisor. Therefore the following is obtained.

$$a_nx_n \equiv b \pmod d,$$

We also have the following

$$\gcd(d, a_n) = (d, a_n) = (a_1, \ldots, a_{n-1}, a_n) = 1,$$

by way of $(a_1, a_2, \ldots, a_n) = 1$. This implies by a prior problem (Diophantine equation $n = 2$) that $\gcd(d, a_n) = 1$ divides b and thus there exists a solution for x_n of $a_nx_n \equiv b \pmod d$. Let this solution be

$$a_nx_n \equiv b \pmod d \Leftrightarrow a_nA \equiv b \pmod d \Leftrightarrow x_n \equiv A \pmod d.$$

The former one implies that there exist a $k \in \mathbb{Z}$ such that $a_nx_n - b = kd$. This further implies that $d | a_nx_n - b$, or $a_nx_n - b$ is a multiple of d or equivalently $(a_nx_n - b)/d = k$ is an integer. All solutions x_n are such that $x_n \equiv A \pmod d$ for $a_nA \equiv b \pmod d$. In equation Eq.(10.44) we substitute as follows noting $d \neq 0$.

$$\begin{aligned}
a_1x_1 + a_2x_2 + \ldots + a_{n-1}x_{n-1} + a_nx_n &= b \\
a_1x_1 + a_2x_2 + \ldots + a_{n-1}x_{n-1} &= -d(a_nx_n - b)/d \\
\frac{a_1}{d}x_1 + \frac{a_2}{d}x_2 + \ldots + \frac{a_{n-1}}{d}x_{n-1} &= -(a_nx_n - b)/d \\
c_1x_1 + c_2x_2 + \ldots + c_{n-1} &= -k,
\end{aligned}$$

where $c_i = a_i/d$. We then have

$$(c_1, c_2, \ldots c_{n-1}) = \frac{1}{d} \cdot (dc_1, dc_2, \ldots dc_{n-1}) = \frac{1}{d} \cdot (a_1, a_2, \ldots a_{n-1}) = \frac{1}{d} \cdot d = 1.$$

By $H(n-1)$ the latter has integer solutions for x_i, $i = 1, \ldots, n-1$ expressed in terms of $n-2$ integer parameters. Adding to it x_n which is $x_n \equiv A \pmod d$ and thus there is an $l \in \mathbb{Z}$ the $n-1$-st integer parameter such that $x_n = A + dl$. ■

Exercise 926.
Solve the equation $3x+5y=7$ on integers.

Solution.
We observer $\gcd(3,5)=1$ and thus by the previous problem it has an integer solution.

$$3x+5y=7 \Leftrightarrow 3x+5y\equiv 7 \pmod 5 \Leftrightarrow 3x\equiv 2 \pmod 5,$$

and from the last equivalence we obtain a solution $x\equiv 4 \pmod 5$ i.e. $x=4+5k$, $k\in\mathbb{Z}$. We go back to $3x+5y=7$ and

$$3x+5y=7 \Leftrightarrow 3(4+5k)+5y=7 \Leftrightarrow y=-1-3k.$$

For $k\in\mathbb{Z}$ the solution space is $x=4+5k$ and $y=-1-3k$. ■

Exercise 927.
Solve the equation $77x - 35y = 21$ on integers.

Solution.
Since $\gcd(77,35) = 7$ divide by 7 to obtain $11x - 5y = 3$. Work out similarly to the previous problem using $(\mathrm{mod}\ 11)$ to obtain $y = 11k - 5$, and then derive an $x = 5k - 2$ to complete the solution. ■

Exercise 928.
Solve the equation $3x+5y+9z=1$ on integers.

Solution.
Since $(3,5,9)=1$, we

$$3x+5y+9z=1 \Leftrightarrow 3x+5y+9z\equiv 1 \pmod 3 \Leftrightarrow 2y\equiv 1 \pmod 3 \Leftrightarrow y\equiv 2 \pmod 3,$$

and therefore $y=3k+2$ for any $k\in\mathbb{Z}$. We then have

$$3x+5y+9z=1 \Leftrightarrow 3x+5(3k+2)+9z=1 \Leftrightarrow 3x+9z=-9-15k \Leftrightarrow 3x+9z\equiv -9-15k \pmod 9 \Leftrightarrow 3x\equiv -6k \pmod 9.$$

The last one has a solution for $x=-2k+9l$ for any $l\in\mathbb{Z}$. Then we again go back to our equation

$$3x+5y+9z=1 \Leftrightarrow 3(-2k+9l)+5(3k+2)+9z=1 \Leftrightarrow z=-1-3l-k.$$

The three x,y,z are expressed in two variables/parameters k,l as needed. ■

Exercise 929.
Show that for $n \in \mathbb{N}$ with $n > 0$ the following applies.

$$\sum_{d|n} \phi(d) = n.$$

Solution.
Consider an integer k among $1,2,\ldots,n$. Let $\gcd(k,n) = d$. Note that $\gcd(k,n) = d$ implies $\gcd(k/d,n/d) = \gcd(m,n/d) = 1$. Let $k/d = m$, or in other words $k = dm$. Then we insert k into set $S_d = \{md : 1 \leq m \leq n/d, \gcd(m,n/d) = 1\}$. The cardinality of S_d is $|S_d| = \phi(n/d)$. Then

$$|\cup_{d:d|n} S_d| = |\{1,2,\ldots,n\}| = n \Leftrightarrow \sum_{d|n} |S_d| = n \Leftrightarrow \sum_{d|n} \phi(n/d) = n \Leftrightarrow \sum_{d|n} \phi(d) = n,$$

sice n/d is a divisor of n if and only if d is a divisor of n. ■

Exercise 930.

Let p be a prime number greather than 1 with $p \equiv 3 \pmod 4$. Show that for all a, b

$$a^2 + b^2 \equiv 0 \pmod p \Longrightarrow a \equiv b \equiv 0 \pmod p.$$

Solution.
We distinguish two cases: (a) $p|a$ and (b) $p \nmid a$. We then show that the latter is impossible, i.e. it is always the case $p|a$.
(a) Let $p|a$. Then $p|a^2$ and therefore $a^2 \equiv 0 \pmod p$ which leads to $a \equiv 0 \pmod p$. If $p|a$ given that $a^2 + b^2 \equiv 0 \pmod p$, and from the former as before we obtain, $p|a^2$ and then $a^2 \equiv 0 \pmod p$, that leads to $b^2 \equiv 0 \pmod p$ i.e. $b \equiv 0 \pmod p$.
(b) Let $p \nmid a$. From

$$\begin{aligned} a^2 + b^2 &\equiv 0 \pmod p \Leftrightarrow \\ a^2 b^{p-3} + b^2 b^{p-3} &\equiv 0 \pmod p \Leftrightarrow \\ a^2 b^{p-3} + b^{p-1} &\equiv 0 \pmod p \Leftrightarrow \\ a^2 b^{p-3} &\equiv -b^{p-1} \pmod p \Leftrightarrow \\ a^2 b^{p-3} &\equiv -1 \pmod p, \end{aligned}$$

where in order to obtain the last step we used Fermat's theorem. Since $p \equiv 3 \pmod 4$, we have $p - 3 = 4k$ for some integer k and $(p-3)/2 = 2k$ is an even integer. For $A = ab^{(p-3)/2}$ we have

$$a^2 b^{p-3} \equiv -1 \pmod p \Leftrightarrow A^2 \equiv -1 \pmod p \Leftrightarrow A^{2 \cdot \frac{(p-1)}{2}} \equiv (-1)^{\frac{(p-1)}{2}} \pmod p \Leftrightarrow A^{p-1} \equiv (-1) \pmod p$$

where we derived the last expression by way of the fact that $(p-1)/2 = 2k+1$ is an odd number. The latter contradicts, by Fermat's theorem, the

$$A^{p-1} \equiv 1 \pmod p,$$

and therefore we can never have $p \nmid a$ of case (b). ■

10.20 Primitive roots

Exercise 931.
Let for $m,m \in \mathbb{Z}$, $\gcd(m,n) = 1$. Show the following

$$a \equiv b \pmod{mn} \Leftrightarrow a \equiv b \pmod{m} \quad \wedge \quad a \equiv b \pmod{n}. \tag{10.45}$$

Solution.
$\Rightarrow$. If $a \equiv b \pmod{mn}$, the there exists an k such that $a - b = kmn$. Since m divides kmn then $m|a-b$. Likewise, $n|a-b$. From the former we have $a \equiv b \pmod{m}$ and the latter $a \equiv b \pmod{n}$.
$\Leftarrow$. Let $a \equiv b \pmod{m} \wedge a \equiv b \pmod{n}$. Then there exist $M,N \in \mathbb{Z}$ such that $a - b = Mm$ and $a - b = Nn$ respectively. Therefore $Mm = Nn$. Since n divides Nn it should divide Mm but because $\gcd(n,m) = 1$ this implies $n|M$. Therefore $M = ln$ for some integer l. Equation $a - b = Mm = lnm = l(nm)$. Therefore $nm|a-b$ and we conclude $a - b \equiv 0 \pmod{mn}$ or equivalently, $a \equiv b \pmod{mn}$.
The result is also a by product of the chinese remainder theorem. ■

Exercise 932.

Let p be a positive prime number, and let g be a primitive root mod p.
Show that $g^i \not\equiv g^j \bmod p$ for all i, j such that $0 \leq i < j < p-1$.

Solution.
Suppose that i, j are such that

$$g^i \equiv g^j \pmod{p}.$$

Then let g^{-1} be the inverse of g mod p, that exists since $\gcd(g, p) = 1$ and we then have the following utilizing the previous assumption.

$$(g)(g^{-1}) \equiv 1 \pmod{p} \Leftrightarrow 1 \equiv (g^i)(g^{-1})^i \Leftrightarrow 1 \equiv (g^j)(g^{-1})^i \Leftrightarrow 1 \equiv g^{j-i}.$$

Then, since $0 < j - i \leq j < p-1$, g can't be a primitive root mod p. ■

Exercise 933.

Let g be a primitive root mod n, for some positive odd integer n. Then show the following.

$$g^{\phi(n)/2} \equiv -1 \pmod{n}.$$

Solution.

Let $n > 2$. Then by Euler's theorem we have the following.

$$g^{\phi(n)} \equiv 1 \pmod{n}.$$

Consider

$$x^2 \equiv 1 \pmod{n}.$$

This implies the following.

$$\begin{aligned} x^2 \equiv 1 \pmod{n} &\Leftrightarrow x^2 - 1 \equiv 0 \pmod{n} \\ &\Leftrightarrow (x-1)(x+1) \equiv 0 \pmod{n} \\ &\Leftrightarrow (x-1) \equiv 0 \pmod{n} \quad \vee \quad (x+1) \equiv 0 \pmod{n} \\ &\Leftrightarrow x \equiv 1 \pmod{n} \quad \vee \quad x \equiv -1 \pmod{n}. \end{aligned}$$

Set $x^2 = g^{\phi(n)}$ and thus the following apply, given that $\phi(n)$ is even and thus $\phi(n)/2$ is an integer.

$$g^{\frac{\phi(n)}{2}} \equiv 1 \pmod{n} \quad \vee \quad g^{\frac{\phi(n)}{2}} \equiv -1 \pmod{n}$$

The former implies that the order of g is no more than $\frac{\phi(n)}{2}$ contradicting the fact that g is a primitive root. Thus the latter applies. ■

Exercise 934.

Let g be a primitive root mod n. Then for $a \in \mathbb{Z}_n$ we have that modular congruence

$$a^2 \equiv 1 \pmod{n},$$

has 2 solutions mod n:

$$a \equiv 1 \pmod{n} \quad \text{and} \quad a \equiv -1 \pmod{n}$$

Solution.
Equation/Congruence

$$a^2 \equiv 1 \pmod{n},$$

implies

$$a^2 \equiv 1 \pmod{n},$$

Thus $a^{-1} = a$ and thus

$$aa^{-1} \equiv 1 \pmod{n}.$$

This means a is a unit and thus $\gcd(a,n) = 1$. This is because if $\gcd(a,n) = d > 1$, then $a^2 \equiv 1 \pmod{n}$ implies $a^2 - 1 = kn$ for some integer k. We have $d|a$, thus $d|a^2$ and we also have $d|n$ and thus $d|kn$ and thus $d|a^2 - kn$ which implies $d|1$ i.e. $d = 1$ contradicting $d > 1$. If $a \in \mathbb{U}_n$, and since g be a primitive root mod n, there exists a k such that $0 < k < \phi(n)$ such that

$$a \equiv g^k \pmod{n} \Leftrightarrow a^2 \equiv g^{2k} \pmod{n} \Leftrightarrow 1 \equiv g^{2k} \pmod{n} \Leftrightarrow g^{2k} \equiv 1 \pmod{n}$$

This leads (see also the previous problem) to the following.

$$g^k \equiv 1 \pmod{n}, k \neq 0 \quad \vee \quad g^k \equiv -1 \pmod{n}, k \neq 0, \quad \vee \quad k = 0.$$

The first is dismissed by g being a primitive root and $k < \phi(n)$. The latter two are the only remaining possibilities. But then

$$a \equiv g^k \equiv -1 \pmod{n},$$

or for $k = 0$,

$$a \equiv g^k \equiv 1 \pmod{n}$$

The result follows. ■

Exercise 935.
Euler's theorem and Fermat's Little theorem are the same for $n = p$ and p a prime number.

Solution.
Consider $p > 1$ a prime number. Then $\phi(p) = p - 1$. Moreover,

$$\mathbb{U}_p = \{1, \ldots, p-1\}$$

Then for any $a \in \mathbb{Z}$, $a > 1$, and $n = p$ with $\gcd(a, n) = 1$ by Euler's theorem we have the following.

$$a^{\phi(n)} \equiv 1 \pmod{n} \Rightarrow a^{\phi(p)} \equiv 1 \pmod{p} \Rightarrow a^{p-1} \equiv 1 \pmod{p},$$

which is Fermat's little theorem. ■

Exercise 936.

Prime power p^k. For prime p and p^k, $k > 1$, the number of units mod p^k and thus the cardinality of $\mathbb{U}_{p^k}$ is p^k minus the multiples of p which is p^{k-1}.

$$|\mathbb{U}_{p^k}| = p^k - p^{k-1} = p^{k-1}(p-1), \quad \phi(p^k) = p^k - p^{k-1}. \tag{10.46}$$

Solution.
The cardinality of $\mathbb{U}_{p^k}$ is $\phi(p^k)$ which is $\phi(p^k) = p^k - p^{k-1}$. ■

Exercise 937.
Let $m,n>1$ be integer. The following two statements are equivalent.

(i) a is a unit mod mn.
(ii) a is a unit mod m $\wedge$ a is a unit mod n.

Solution.
$(i)\Rightarrow(ii)$. If a is a unit $\pmod{mn}$ then it means $\gcd(a,mn)=1$.
We claim that this implies that $\gcd(a,m)=\gcd(a,n)=1$.
If this was not so, and say $\gcd(a,m)=d>1$, then d becomes a common divisor of a and m (and consequently of mn as well). This would imply $\gcd(a,mn)>1$, a contradiction to $\gcd(a,mn)=1$.
Then $\gcd(a,m)=\gcd(a,n)=1$. Therefore $ax\equiv 1 \pmod{m}$ has a solution for x and thus a is unit mod m. Likewise, $ay\equiv 1 \pmod{n}$ has a solution for y and thus a is unit mod n. Case completed.
$(ii)\Rightarrow(i)$. If a is a unit mod m then $\gcd(a,m)=1$. If a is a unit mod n then $\gcd(a,n)=1$. The former imply $\gcd(a,mn)=1$.
If the latter was not so, then $\gcd(a,mn)=d>1$. There is a prime factor p of d i.e. $p|d$. Then $p|a$ and thus $p|mn$. The latter implies $p|m$ or $p|n$. One or the other combined with $p|a$ implies that $p|\gcd(a,m)$ or $p|\gcd(a,n)$ contradicting that $\gcd(a,m)=\gcd(a,n)=1$. The result then follows. ■

Exercise 938.

Let $n \in \mathbb{N}$ and let $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$. Let $ord_n(a) = k$, that is k is the smallest positive integer such that $a^k \equiv 1 \pmod{n}$.
(a) Show that $k|\phi(n)$ or in other words $ord_n(a)|\phi(n)$.
(b) Moreover, show that for all $b,c \in \mathbb{N}$, and a such that $\gcd(a,n) = 1$, the following applies

$$a^b \equiv a^c \pmod{n} \Longleftrightarrow b \equiv c \pmod{k}. \tag{10.47}$$

(c) After (b) show that for a such that $\gcd(a,n) = 1$, the following applies.

$$a^b \equiv 1 \pmod{n} \Longleftrightarrow k|b, \tag{10.48}$$

where $k = ord_n(a)$.

Solution.
(a) It follows directly from part (c) using part (b). Since by Euler's theorem

$$a^{\phi(n)} \equiv 1 \pmod{n},$$

part (c), with $b = \phi(n)$ concludes that $k|\phi(n)$.
A more direct approach to proving part (a) follows. Consider minimum $k = ord_n(a)$ such that

$$a^k \equiv 1 \pmod{n},$$

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

We have $k \leq \phi(n)$. Then let $\phi(n) = Ak + R$, where $0 \leq R < k = ord_n(a)$. We have the following

$$a^{\phi(n)} \equiv 1 \pmod{n} \Leftrightarrow a^{Ak+R} \equiv 1 \pmod{n} \Leftrightarrow (a^k)^A a^R \equiv 1 \pmod{n} \Leftrightarrow a^R \equiv 1 \pmod{n},$$

and the last one if $R > 0$ then $R < k = ord_n(a)$, a contradiction to the minimality of k. Thus $R = 0$ is the only possibility leading to $\phi(n) = Ak$ and thus $ord_n(a)|\phi(n)$.
(b) $\Leftarrow$. If $b < c$ we swap, and thus in the remainder $b \geq c$. If $b \equiv c \pmod{n}$ then there exists an l such that $b - c = kl$. Since $b \geq c$, then $l \geq 0$. We then obtain the following.

$$a^b = a^{kl+c} = a^c (a^k)^l \equiv a^c \cdot 1 \pmod{n}.$$

This implies the following.

$$a^b \equiv a^c \pmod{n},$$

and the result is proven.
$\Rightarrow$. Let $a^b \equiv a^c \pmod{n}$. Then the following are applicable.

$$\begin{aligned} a^b \equiv a^c \pmod{n} &\Leftrightarrow a^b - a^c \equiv 0 \pmod{n} \Leftrightarrow \\ (a^{b-c} - 1)a^c \equiv 0 \pmod{n} &\Leftrightarrow (a^{b-c} - 1) \equiv 0 \pmod{n} \Leftrightarrow \\ a^{b-c} \equiv 1 \pmod{n}, & \end{aligned}$$

where the last two equivalences is due to the fact $\gcd(a,n) = 1$ and thus $n \nmid a^c$. Consider $b - c$. Then $b - c = kl + m$, where $0 \leq m < k$. Then the following are applicable.

$$\begin{aligned} a^{b-c} \equiv 1 \pmod{n} &\Leftrightarrow a^{kl+m} \equiv 1 \pmod{n} \Leftrightarrow \\ a^{kl} a^m \equiv 1 \pmod{n} &\Leftrightarrow (a^k)^l a^m \equiv 1 \pmod{n} \Leftrightarrow \\ a^m \equiv 1 \pmod{n}. & \end{aligned}$$

The last inequality derives from the fact k is the smallest positive integer $a^k \equiv 1 \pmod{n}$. Note that m can be 0 or $1,\ldots,k-1$. It cannot be $1,\ldots,k-1$ because the smallest m with $a^m \equiv 1 \pmod{n}$ that is a positive integer is k and nothing smaller such as m. Thus the only other possibility is for $m=0$. The $b-c=kl+0=kl$ therefore $k|b-c$ i.e

$$b-c \equiv 0 \pmod{k} \Leftrightarrow b \equiv c \pmod{k},$$

as needed. This completes part (b).
(c) The choice of k is such that $a^k \equiv 1 \pmod{n}$ and this is the smallest positive integer. Consider a b such that $a^b \equiv 1 \pmod{n}$. Then $1 \equiv a^b \equiv a^k \pmod{n}$ and by part (b), we have $b \equiv k \pmod{k}$. This implies $b-k \equiv 0 \pmod{k}$ and therefore $k|b$.
Alternative proof for (c). Let $b=kk_1+r$, where $0 \leq k < r$. Then

$$a^b \equiv a^{kk_1+r} \equiv (a^k)^{k_1}a^r \equiv a^r \pmod{n}.$$

Note that in $a^b \equiv a^r \pmod{n}$, we have $r<k$. If $a^b \equiv 1 \pmod{n}$ then $a^r \equiv 1 \pmod{n}$, and we have found an $0<r<k$, an impossibility to the positive minimality of k. Otherwise $r=0$ but then $b=kk_1$ which implies $k|b$ as needed and we are done. ■

Exercise 939.
Let $\text{ord}_n(a) = k$ and $\text{ord}_n(b) = l$. If $\gcd(k,l) = 1$ then $\text{ord}_n(ab) = kl$. Therefore show the following.

$$\text{ord}_n(a) = k \quad \wedge \quad \text{ord}_n(b) = l \quad \wedge \quad \gcd(k,l) = 1 \Rightarrow \text{ord}_n(ab) = kl. \tag{10.49}$$

Solution.
Let $\text{ord}_n(ab) = m$. Then

$$(ab)^{kl} \equiv (a^k)^l (b^l)^k \equiv 1 \pmod{n}$$

Therefore $m|kl$. Moreover

$$1 \equiv ((ab)^m)^k = (a^k)^m (b^{km}) \equiv b^{km} \pmod{n}$$

This means $l|km$. Since $\gcd(k,l) = 1$ we have $l|m$. Likewise, $k|m$. Since $\gcd(k,l) = 1$ we have $kl|m$. This implies $kl \leq m$. But since m is the order of ab we must also have, by the first derivation above, $kl \geq m$. Thus $kl = m$. ■

Exercise 940.
Let $n \in \mathbb{Z}_+$ and there exists a primitive root mod n. Then $\phi(\phi(n))$ of the $1, 2, \ldots, n-1$ are primitive roots mod n.

Solution.
If a is a primitive root mod n, then $a \bmod n$ is a generator of $(\mathbb{Z}/n\mathbb{Z})^x$. We thus need to find b^i such that $1 \leq i \leq \phi(n)$ where $a^i \pmod n$ is a generator of $(\mathbb{Z}/n\mathbb{Z})^x$. For a^i to be a generator we need to have $(a^i)^j \equiv a^k \pmod n$, for every $1 \leq k \leq \phi(n)$. In other words $ij \equiv k \pmod{\phi(n)}$. The latter is true for $\gcd(i, \phi(n))|k$, for all i or $\gcd(i, \phi(n)) = 1$. The number of elements i is just $\phi(\phi(n))$. ■

Exercise 941.
Find the primitive roots mod 7.

Solution.
Since $\phi(7) = 6$, we consider a such that $a^6 \equiv 1 \pmod 7$. Consider $a = 1$, is such. $a = 3$ has $3^1, 3^2, 3^3$ equiv to $3, 2, 6$ respectively. Moreover $3^4, 3^5, 3^6$ are equiv to $4, 5, 1$ respectively. Thus 3 as order 6. For 3^i to be a generator by the previous problem $\gcd(i, \phi(7)) = 1$. The only i such that $\gcd(i, 6) = 1$ are 1,5. So indeed 3 is one such generator. The other one is $3^5 \equiv 5 \pmod 7$. Note that since $\phi(\phi(7)) = \phi(6) = 2$, there is no other one. ■

Exercise 942.

(a) Show that if $\text{ord}_n(a) = k$, then a^m has order k if and only if $\gcd(m,k) = 1$.

$$\text{ord}_n(a) = k : ord_n(a^m) = k \Leftrightarrow \gcd(m,k) = 1. \tag{10.50}$$

(b) Show that if $\text{ord}_n(g) = \phi(n) = k$, then g^m has order $\phi(n)$ if and only if $\gcd(m,\phi(n)) = 1$.

$$\text{ord}_n(g) = k = \phi(n) : ord_n(g^m) = \phi(n) \Leftrightarrow \gcd(m,\phi(n)) = 1. \tag{10.51}$$

(c) Show that if $\mathbb{Z}_n$ has a primitive root g, then all the primitive roots of $\mathbb{Z}_n$ are those g^m such that $\gcd(m,\phi(n)) = 1$. In particular, there are $\phi(\phi(n))$ primitive roots mod n.

$$\exists g \in \mathbb{U}_n : ord_n(g) = \phi(n) \Rightarrow |\{x : x \text{ primitive root mod } n\}| = |\{g^m : ord_n(g^m) = \phi(n)\}| = \phi(\phi(n)). \tag{10.52}$$

Solution.
(a) Let $d = \gcd(m,k)$ and let $q = \text{ord}_n(a^m)$. We need to show that $q = k$ if and only if $d = 1$.
$\Rightarrow$.
Let $q = k$. We will show that $d = 1$. Let $d > 1$. Then $k = dx$ and $m = dy$, for some integer x, y. Note that $x < k$ and $y < m$. Then we have

$$(a^m)^x \equiv a^{mx} \equiv a^{dyx} \equiv (a^k)^y \equiv 1 \pmod{p}$$

Thus order q of a^m is x. Since $x < k = q$ the element a^m has "order" x less than k less than its order $q = k$. It is impossible (or contradiction to the minimality of $q = k$). Thus it should be $x = k$ which implies $d = 1$. Result shown.
$\Leftarrow$.
If $d = 1$ we show that $q = k$. Since $q = \text{ord}_n(a^m)$ we have

$$a^{mq} \equiv (a^m)^q \equiv 1 \pmod{p}$$

This means that $k|mq$ from Eq. (10.48). Since $d = \gcd(k,m) = 1$, we have $k|q$. Thus $k \leq q$. We also have

$$(a^m)^k \equiv (a^k)^m \equiv 1 \pmod{p}$$

Thus $q|k$ i.e. $q \leq k$. Combining the latter with a previous $k \leq q$ the result follows.
(b) In part (a) pick $a = g$ and $ord_n(a) = ord_n(g) = k = \phi(n)$.
(c) If g is a primitive root then $\text{ord}_n(g) = \phi(n)$ by definition. By part (a) or (b), $\text{ord}_n(g^m) = \phi(n)$ if and only if $\gcd(m,\phi(n)) = 1$.
Moreover, the number of values m such that this is true is $\phi(\phi(n))$. All elements of $\mathbb{U}_n$ are of the form g^i or $(g^m)^i$ and thus the primitive roots will thus be found this way. Note that $|\mathbb{U}_n| = \phi(n)$, but the number of primitive roots mod n is $\phi(\phi(n))$.
Note. Let g be a primitive root mod n with $\gcd(g,n) = 1$. Then set

$$\mathbb{U}_n = \{g^1, g^2, \ldots, g^{\phi(n)} \bmod$$

noting by Euler's theorem that the last term is $g^{\phi(n)} \equiv 1 \pmod{n}$. Since $\gcd(g,n) = 1$ then $\gcd(g^i,n) = 1$ for all $i = 1, \ldots, \phi(n)$. Moreover $g^i \not\equiv g^j \pmod{n}$, $i \neq j$ thus implying $|\mathbb{U}_n| = \phi(n)$, where $1 \leq i, j \leq \phi(n)$. To prove $g^i \not\equiv g^j \pmod{n}$ let us assume that $g^i \equiv g^j \pmod{n}$, reaching a $g^{i-j} \equiv 1 \pmod{n}$, where $0 \leq i - j < \phi(n)$. Since g is a primitive root $\phi(n)|i-j$ i.e. $\phi(n) \leq i - j$ which contradicts $i - j < \phi(n)$. Thus the only possibility is that $i - j = 0$ i.e. $i = j$ which contradicts the hypothesis $i \neq j$, since otherwise we would have derived $ord_n(g) = i - j < \phi(n)$, a contradiction! ■

Exercise 943.

Let p be a prime number, and g a primitive root mod p and therefore $ord_p(g) = \phi(p) = p-1 = k$. Then g^m has

$$ord_p(g^m) = \frac{p-1}{\gcd(m, p-1)}. \tag{10.53}$$

Solution.
Let $d = \gcd(m, p-1)$. Then there exists integers a, b such that

$$m = da, \quad p-1 = db, \quad \gcd(a,b) = \gcd(\frac{m}{d}, \frac{p-1}{d}) = 1.$$

Let $ord_p(g^m) = l$. Consider $(g^m)^l$.

$$g^{ml} \equiv (g^m)^l \equiv 1 \pmod{p},$$

by defition of $ord_p(g^m) = l$. Then $ord_p(g)|ml$ that is, $p-1|ml$. But then $(p-1)/d|(m/d)l$, and since $\gcd(\frac{m}{d}, \frac{p-1}{d}) = 1$ we conclude that $((p-1)/d|l$ or $(p-1)/d \leq l$.
Furthemore consider $(g^m)^{\frac{p-1}{d}}$.

$$(g^m)^{\frac{p-1}{d}} \equiv (g)^{m\frac{p-1}{d}} \equiv (g)^{da\frac{p-1}{d}} \equiv (g)^{(p-1)a} \equiv (g^{p-1})^a \equiv 1 \pmod{p},$$

since $ord_p(g) = p-1$. The latter implies $ord_p(g^m)|\frac{p-1}{d}$ and thus $l = ord_p(g^m) \leq (p-1)/d$. From the prior $(p-1)/d \leq l$ and the current $l = ord_p(g^m) \leq (p-1)/d$ we conclude $l = ord_p(g^m) = (p-1)/d$, as needed. ■

Exercise 944.

Let p,q be prime numbers $p \neq q$. Show that $\mathbb{Z}_{pq}$ has no primitive root.

Solution.
It is the case that $\gcd(p,q) = 1$ by $p \neq q$, for prime p,q. Pick an element x of $\mathbb{Z}_{pq}^x = \mathbb{U}_{pq}$. We have by Euler's theorem

$$x^{\phi(pq)} \equiv 1 \pmod{pq} \Leftrightarrow x^{\phi(p)\phi(q)} \equiv 1 \pmod{pq}$$

Note that $\phi(pq) = \phi(p)\phi(q)$ for prime p,q. It is easy to show that $\phi(p)$ and $\phi(q)$ are even and thus $\phi(p)/2$ and $\phi(q)/2$ are integer. Consider

$$k = \frac{\phi(p)\phi(q)}{2}.$$

We have $\phi(p)|k$ and $\phi(q)|k$. Let us assume that there exists a primitive root g mod pq. Then $ord_{pq}g = \phi(pq) = \phi(p)\phi(q)$. Moreover,

$$g^k \equiv 1 \mod p \quad , \quad g^k \equiv 1 \mod q,$$

since for example

$$g^k \equiv g^{\frac{\phi(p)\phi(q)}{2}} \equiv (g^{\phi(p)})^{\frac{phi(q)}{2}} \equiv 1 \pmod{p},$$

with a similar and symmetric proof for a mod q result. Then

$$g^k \equiv 1 \pmod{pq},$$

and given that $k < \phi(pq)$ since $k = \frac{\phi(p)\phi(q)}{2}$, we have a contradiction to the assumption that g is a primitive root mod pq.

■

Exercise 945.

Let $(S,+,\cdot)$ be a field and let $f \in S[x]$, where $f \neq 0$ be a polynomial over S. Then f has at most deg(f) roots.

Solution.
A proof follows by induction.
Base case. Let $n = deg(f) = 0$ be the degree of f, that is, $f(x) = a_0$. Since $f(x) \neq 0$ we have $a_0 \neq 0$ and thus the polynomial has at most 0 roots that is 0 roots obviously. The base case has been proven.
Inductive step. Assume that the claim holds for polynomials of degree up to $n-1$. Consider $n = deg(f) \geq 1$. If polynomial f has at most n roots, then we are done. Consider $f(x)$ as follows.

$$f(x) = a_n x^n + \ldots + a_1 x + a_0 \in S[x],$$

with $f(x)$ being a polynomial of degree n. If $f(x)$ has no roots in S, then we are also done. Otherwise assume that $f(x)$ has at least one root, and let it be r that is $f(r) = 0 \pmod{n}$. Then we have the following.

$$f(x) = f(x) - f(a) = a_n(x^n - r^n) + \ldots + a_1(x-r) = \sum_{i=1}^{n} a_i(x^i - r^i).$$

We have that $x^i - r^i = (x-r)\sum_{j=0}^{i-1} x^j r^{i-1-j}$. Then

$$f(x) = (x-r)g(x),$$

where $deg(g) \leq n-1$. Consider another root of f and let it be s. For $s \neq r$, we have $0 = f(s) = (s-r)g(s)$. Since $s \neq r$ this implies $g(s) = 0$. This is because element $s - r \neq 0$ has an inverse and multiplying both sides of $0 = (s-r)g(s)$ with that inverse we conclude that $g(s) = 0$ that is s is a root of g. By the induction hypothesis applied to g there can be no more than $n-1$ roots of g such as s. Adding to this root r we can have at most n roots for f.
Consider $Z_p = \mathbb{Z}/p\mathbb{Z}$, where p is a prime number. Then $(Z_p,+,\cdot)$ is a field. Thus the following result due to Lagrange is applicable. See the next problem, already proven for the more general case of a field in this problem. ■

Exercise 946.
If $x^{p-1}-1=f(x)g(x)$ and $deg(f)=k$ and $deg(g)=l$ then $f(x)$ has k distinct roots $\pmod p$ and $g(x)$ has l.

Solution.
Note that $p-1=k+l$. If t is a root of $x^{p-1}-1$ then $f(t)g(t)\equiv 0 \pmod p$. Thus either $f(t)$ or $g(t)$ is $\equiv 0 \pmod p$. If one has fewer than the alotted roots then the other has more than the alotted thus if f has $<k$ then g has $>l$. ■

Exercise 947.

Find the primitive roots mod 11 and mod 10..

Solution.

Let us start with $n = 11$. Since $\phi(11) = 10$, we consider a such that $a^{10} \equiv 1 \pmod{1}1$. Consider $a = 2$. We do have $2^{10} \equiv 1 \pmod{1}1$. $a = 2$ has $2^1, 2^2, 2^3$ equiv. to $2, 4, 8$ respectively. Moreover $2^4, 2^5, 2^6$ are equiv. to $5, 10, 9$ respectively. Furthermore, $2^7, 2^8, 2^9$ are equiv. to $3, 6, 1$ respectively. Thus 2 has order 10, as non of those powers from 1 to 9 are 1 mod 11; the only one is 10 for $2^{10} \equiv 1 \pmod{1}1$. For 2^i to be a generator by a prior problem $\gcd(i, \phi(11)) = 1$. The only i such that $\gcd(i, 10) = 1$ are $i = 1, 3, 7, 9$. So indeed $2^1 = 2$ is one such generator. The other ones are $2^3 \equiv 8 \pmod{1}1$, $2^7 \equiv 7 \pmod{1}1$, $2^9 \equiv 6 \pmod{1}1$. Note that since $\phi(\phi(11)) = \phi(10) = 4$, there is no other one.

Let us continue with $n = 10$. Since $\phi(10) = \phi(5 \cdot 2) = \phi(2) \cdot \phi(5) = 4$, we consider an a such that $a^{\phi(10)} \equiv 1 \pmod{1}0$. Consider $a = 3$. We do have $3^4 \equiv 1 \pmod{1}0$. $a = 3$ has $3^1, 3^2, 3^3$ equiv. to $3, 9, 7$ respectively. Moreover $3^4, 3^5, 3^6$ are equiv. to $1, 3, 9$ respectively. Furthermore, $3^7, 3^8, 3^9$ are equiv. to $7, 1, 3$ respectively. Thus 3 has order 4. For 3^i to be a generator by a prior problem $\gcd(i, \phi(10)) = 1$. The only i such that $\gcd(i, 4) = 1$ are $i = 1, 3$. So indeed $3^1 = 3$ is one such generator. The other ones are $3^3 \equiv 7 \pmod{1}0$. Note that since $\phi(\phi(10)) = \phi(4) = 2$, there is no other one. ■

10.20.1 Lagrange's theorem

Exercise 948.

Lagrange's theorem. Consider a polynomial $f(x)$ of degree n with integer coefficients. Let p be a prime number.

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0.$$

Consider $a_n \bmod p \neq 0$. That is we exclude from consideration the case that all coefficients are multiples of p or equivalently exclude all cases $p|a_i$ for all $i = 0,1,\ldots,n$. Then $f(x)$ can have at most n distinct zeroes mod p.

In other words, $f(x) \equiv 0 \pmod{p}$, has at most $deg(f) = n$ roots mod p, unless all (presumed integer) coefficients of f are divisible by p.

Note 1. No more than n elements of $\mathbb{U}_p$ can thus be a root of $f(x) \equiv 0 \pmod{p}$.

Note 2. The "expression n zeroes mod p" is to be read "n congruence classes of solutions (zeroes) mod p". Thus if k, an elements of $\mathbb{U}_p$ is a root of $f(x) \equiv 0 \pmod{p}$ and $f(x) \equiv 0 \pmod{p}$ so do we expect for $k+p, k+2p, \ldots$.

Solution.

A proof follows by induction.

Base case. For $n = deg(f) = 0$ we have $f(x) = a_0$ that has 0 roots mod p for $f(x) \equiv a_0 \equiv 0 \pmod{p}$, for $p \nmid a_0$. (Or for $n = 0$, $a_n \bmod p \neq 0$.)

Inductive step. Consider $n = deg(f) \geq 1$. If f has at most n roots we are done. By the inductive step f has at least one root and let it be r. Consider the polynomials

$$f(x) = g(x)(x-r) + R(x),$$

with $deg(R) < deg(x-r) = 1$ i.e. $deg(R) = 0$. We can write $R(x) = c$ for some integer c. Set $x = r$ and we have

$$(f(r) \equiv 0 \pmod{p}) \equiv g(r) \cdot 0 + c \pmod{p}$$

Then

$$c \equiv 0 \pmod{p}.$$

This means $R(x) = R = c = 0 \pmod{p}$ and therefore

$$f(x) \equiv g(x)(x-r) \pmod{p}.$$

Then, $deg(x-r) = 1$ and $deg(g) \leq n-1$. By the inductive hypothesis the former has one root (and it is r) and the latter has at most $n-1$ roots mod p. Combined $f(x)$ has at most n roots mod p, as needed. ■

Exercise 949.
(a) Consider the polynomial

$$f(x) = x^2 - 1,$$

where $f(x) \in (\mathbb{Z}/9\mathbb{Z})[x]$. Find its roots.
(b) Consider the polynomial

$$f(x) = x^2 - 1,$$

where $f(x) \in (\mathbb{Z}/8\mathbb{Z})[x]$. Find its roots.

Solution.
(a) They are $x = 1, 8 \pmod 9$.
(b) They are $x = 1, 3, 5, 7 \pmod 8$. ■

Exercise 950.

Show the following. If p is a prime number and d is a divisor of $p-1$, then the polynomial $f(x) = x^d - 1 \in Z_p[x]$ has exactly d roots.

Solution.
Since $d|p-1$ then $p-1 = dk$. Consider $x^{p-1}-1$. By Euler's (or FLT) all non zero elements $1,\ldots,p-1$ are roots of it. Therefore it has $p-1$ different roots. Moreover

$$(x^d)^k - 1 = (x^d - 1)((x^d)^{k-1} + \ldots + (x^d)^0) = f(x)g(x),$$

where $g(x)$ is a polynomial of degree $dk - d = p - 1 - d$. By the previous result f has at most d roots, and g at most $dk - d = p - 1 - d$. But the product has at least $p-1$ roots all of them distinct. Thus $f(x)$ has exactly d and $g(x)$ $p-1-d$ roots. The result follows. ■

Exercise 951.

Consider the polynomial $f(x) = x^{p-1} - 1$, where p is a prime number. By Fermat's Little theorem we know. that

$$f(x) \equiv 0 \pmod{p} \quad \Leftrightarrow \quad x^{p-1} \equiv 1 \pmod{p},$$

for any x such that $\gcd(x, p) = 1$. There are $p-1$ such x and these are the elements of $\mathbb{U}_p = \{1, 2, \ldots, p-1\}$ since $|\mathbb{U}_p| = \phi(p) = p-1$. Lagrange's theorem confirms that there are at most $p-1$ distinct solutions (roots) for $f(x) \equiv 0 \pmod{p}$.

Corollary of Lagrange's theorem. Let $f(x) = x^{p-1} - 1$, and let p be a prime number. Consider d a positive integer such that $d|p-1$. Then the modular equation below has at most d distinct solutions mod p

$$x^d \equiv 1 \pmod{p} \qquad \text{has at most d solutions mod p for,} \quad d|p-1. \tag{10.54}$$

Solution.
Since $d|p-1$ then let $k = (p-1)/d$. We observe the following

$$x^{p-1} - 1 = x^{dk} - 1 = (x^d - 1)(x^{(k-1)d} + \ldots + 1).$$

Considering $\pmod{p}$ of both sides we conclude

$$x^{p-1} - 1 \equiv (x^d - 1)(x^{(k-1)d} + \ldots + 1) \equiv 0 \pmod{p}.$$

By the prior discussion $x^{p-1} - 1 \equiv 0 \pmod{p}$ has $p-1$ distinct roots mod p. Say that $x^d - 1 \equiv 0 \pmod{p}$ has LESS than d solutions (roots). By Lagrange's theorem $x^{(k-1)d} + \ldots + 1) \equiv 0 \pmod{p}$ has at most $(k-1)d$ distinct solutions mod p. Thus the product $(x^d - 1)(x^{(k-1)d} + \ldots + 1) \equiv 0 \pmod{p}$ is going to have a number of solutions equal to the sum of (less than d) for $x^d - 1 \equiv 0 \pmod{p}$ and at most $(k-1)d$ for the second term a total of less than $d + (k-1)d = kd = p-1$. This "less than" contradicts the fact that $x^{p-1} - 1 \equiv 0 \pmod{p}$ has exactly $p-1$ solutions mod p. ■

Exercise 952.

Let p be a prime number. Let

$$x^n \equiv 1 \pmod{p},$$

has n distinct roots mod p. Then $\mathbb{U}_p = \mathbb{Z}_p^x$ has exactly $\phi(n)$ elements of order n.
A corollary to this is that

$$x^{p-1} \equiv 1 \pmod{p},$$

has $\phi(p-1) = \phi(\phi(p))$ elements of order $p-1$ i.e. that number of primitive roots mod p.

Solution.
Proof is by induction.
Base case $n = 1$**.** If $x \equiv 1 \pmod{p}$ has one solution mod p, and indeed the only solution is $x = 1$, then $\mathbb{U}_p$ has exactly $\phi(1) = 1$ element of order $n = 1$ and this is 1 mod p. Case completed.
Inductive step. (Strong) Induction that is, assume that result is true for all exponents $< n$ and we show the result for n. That is if $x^k - 1$ has exactly k distinct roots mod p, for all $k < n$ then $\mathbb{U}_k$ has exactly $\phi(k)$ elements of order k. Consider the case for n.
Let d be a divisor of n i.e. there exists integer N such that $n = dN$. Just like in the previous problem we obtain the following.

$$x^n - 1 = (x^d - 1)(x^{d(N-1)} + x^{d(N-2)} + \ldots + x^d + 1).$$

By Lagrange's theorem the last polynomial has at most $d(N-1)$ distinct roots mod p. The left-most polynomial has at most n by Lagrange's theorem. But we are told that $x^n - 1$ has exactly n distinct roots mod p. Then $x^d - 1$ cannot have fewer than d roots since then

$$(= n) \neq (< n) = (< d) + (\leq d(N-1)),$$

$x^n - 1$ would have fewer than n distinct roots. We note also that $d < n$. Thus $x^d - 1$ must have exactly d distinct solutions (roots). By the induction hypothesis since $d < n$ it has among the exactly d distinct roots $\phi(d)$ of order d. Every root of $x^n - 1$ is a root of $x^d - 1$ or the rest of the polynomial (the other factor). Among the n roots of $x^n - 1$, d of them are roots of $x^d - 1$ and distinct; moreover among the d latter roots $\phi(d)$ are of order d mod p. Integer d is an arbitrary divisor of n. One such divisor is n itself, and all the other divisors are less than n. We know then the following

$$n = \sum_{d|n} \phi(d) = \sum_{d|n, d\neq n} \phi(d) + \sum_{d|n, d=n} \phi(d) = \sum_{d|n, d\neq n} \phi(d) + \phi(n).$$

Moreover among the n roots of $x^n - 1$, some of them are of order n, and let that number be $a(n)$ and the rest are of order $<$ than n. We concluded earlier that for every divisor d of n there are $\phi(d)$ roots of n of order d. Adding up all of them we come up with the following.

$$(\text{NumRoots of order } < n) = \sum_{d\neq n, d|n} = \phi(d),$$

and we also have the following.

$$(\text{NumRoots of order } < n) + (\text{NumRoots of order } = n) = (\text{NumRoots of order } < n) + a(n) = n$$

It is obvious that

$$(\text{NumRoots of order } = n) = a(n) = n - \sum_{d\neq n, d|n} \phi(d) = \phi(n).$$

■

Exercise 953.

A result similar to the previous one. Let p be a prime number. Let there exists an a such that $ord_p(a) = k$. Then the number of elements mod p of order k is $\phi(k)$.

Solution.
Since $ord_p(a) = k$, we have

$$a^k \equiv 1 \pmod{p}$$

Consider the sequence

$$a^0, a^1, \ldots, a^{k-1}$$

are the k roots of $x^k - 1 \equiv 0 pmod p$, or equivalently $x^k \equiv 1 \pmod{p}$. This is because for $i = 0, 1, \ldots, k-1$, we have the following.

$$(a^i)^k - 1 \equiv (a^k)^i - 1 \equiv 1 - 1 \equiv 0 \pmod{p}.$$

We now ask the question: Out of the sequence $a^0, a^1, \ldots, a^{k-1}$ how many of them are of order k? The answer comes by way of Eq.(10.53) that states

$$ord_p(a^i) = \frac{ord_p(a)}{\gcd(i, ord_p(a))} = \frac{k}{\gcd(i,k)}.$$

Thus

$$ord_p(a^i) = \frac{k}{\gcd(i,k)} = k,$$

for all i such that $\gcd(i,k) = 1$ and thus for $\phi(k)$ values i. ■

10.20.2 Wilson's theorem revisited

Exercise 954.
A proof of Wilson's theorem using primitive roots. Let p be a prime number then we obtain the following.

$$(p-1)! \equiv -1 \pmod{p}.$$

Solution.
For $p = 2$, it is obvious by direct inspection.
For $p > 2$, then prime number p is odd. Then there exists a primitive root mod p and let it be g. Then the sequence

$$g^1, g^2, \ldots, g^{p-1},$$

is a permuation of

$$1, 2, \ldots, p-1.$$

Then multiplying the elements of the sequence and equating the products mod p we obtain the following.

$$(p-1)! \equiv g^{1+2+\ldots+(p-1)} \equiv g^{\frac{p(p-1)}{2}} \equiv (g^{\frac{(p-1)}{2}})^p \pmod{p}.$$

Given that x is a primitive root, we have that

$$g^{\frac{(p-1)}{2}} \not\equiv 1 \pmod{p},$$

since otherwise we would have obtained an order of g less that $p-1$. However it is also the case that

$$g^{p-1} \equiv 1 \pmod{p},$$

which implies the following.

$$g^{\frac{(p-1)}{2}} \equiv -1 \pmod{p},$$

which implies then the following.

$$g^{\frac{p(p-1)}{2}} \equiv (-1)^p \pmod{p},$$

which results, given that p is an odd prime number to the following.

$$g^{\frac{p(p-1)}{2}} \equiv -1 \pmod{p}.$$

■

10.20.3 Primitive roots continued

Exercise 955.
Let $a \in \mathbb{U}_p = \mathbb{Z}_p^x$, where p is an odd prime number and positive. Then we have the following

$$ord_{p^2}(a) = ord_p(a) \quad \vee \quad ord_{p^2}(a) = p \cdot ord_p(a). \tag{10.55}$$

Solution.
Let

$$ord_p(a) = k \quad \Rightarrow \quad a^k \equiv 1 \pmod{p} \quad , \tag{10.56}$$
$$ord_{p^2}(a) = l \quad \Rightarrow \quad a^l \equiv 1 \mod p^2. \tag{10.57}$$

By way of Eq.(10.56) we have that there exists A such that

$$\begin{aligned} a^k - 1 = Ap \quad &\Rightarrow \quad a^{kp} = (1+Ap)^p \equiv 1 \pmod{p^2} \\ &\Rightarrow \quad ord_{p^2}(a) | kp \\ &\Rightarrow \quad l | kp. \end{aligned} \tag{10.58}$$

Furthermore,

$$\begin{aligned} a^l - 1 = Bp^2 \quad &\Rightarrow \quad a^l \equiv 1 \pmod{p} \\ &\Rightarrow \quad ord_p(a) | l \\ &\Rightarrow \quad k | l. \end{aligned} \tag{10.59}$$

By way of Eq.(10.59) we have $l = Kk$ for some integer K. By way of Eq.(10.58) we have

$$l | kp \Rightarrow Kk | kp \Rightarrow K | p \Rightarrow K = 1 \quad \vee K = p,$$

given that p is a prime number. Since $l = Kk$ this leads to either $l = k$ or $l = pk$ and the result is proven. ■

Exercise 956.
Let p be a prime number. Show that for $p-1 \nmid k$

$$A_k = 1^k + 2^k + \ldots + (p-1)^k,$$

is divisible by p and thus $A_k \equiv 0 \pmod{p}$.

Solution.
Consider g a primitive root mod p. Then

$$g^{p-1} \equiv 1 \pmod{p}.$$

Given that $p-1 \nmid k$ we form the sum B_k as follows. Note that $g^0, g^1, \ldots, g^{p-2}$ is a permutation of $1, 2, \ldots, p-1$ and thus $B_k = A_k$.

$$B_k = g^{0\cdot k} + g^{1\cdot k} + \ldots + \ldots g^{(p-2)\cdot k} = \frac{g^{p\cdot k}-1}{g^k-1} \pmod{p} = \frac{1-1}{g^k-1} \pmod{p} = 0 \pmod{p}$$

For the latter part we used the fact that $p-1 \nmid k$ and thus $g^k - 1 \not\equiv 0 \pmod{p}$. (Otherwise $g^k \equiv 1 \pmod{p}$ implying $ord_p(g) | k = p-1 | k$.) Therefore the geometric sequence has a closed form answer that is 0 mod p. The conclusion is that $p | B_k = p | A_k$ that is A_k is divisible by p. ■

Exercise 957.

Let p be a prime number. Let q be another prime number such that $q^r|p-1$ for some positive integer r. Then show that there exists a $g \in \mathbb{U}_p$ such that

$$ord_p(g) = q^r. \tag{10.60}$$

Solution.
Let us assume that there exists a g such that

$$g^{q^r} \equiv 1 \pmod{p}. \tag{10.61}$$

Then

$$ord_p(g)|q^r.$$

Then, let $ord_p(g) = q^k$ for some $0 < k \leq r$. We obtain the following.

$$g^{q^k} \equiv 1 \pmod{p} \Leftrightarrow g^{q^j} \equiv 1 \pmod{p}, \quad \text{for} \quad j \geq k.$$

The latter is because

$$g^{q^j} \equiv g^{q^{j-k+k}} \equiv (g^{qk})^{q^{j-k}} \equiv (1)^{q^{j-k}} \equiv 1 \pmod{p}.$$

If $ord_p(g) = q^k$ for some $0 < k < r$ (note that it is $k < r$ not $k \leq r$), then

$$ord_p(g) = q^k \Rightarrow g^{q^{r-1}} \equiv 1 \pmod{p} \quad \text{for} k < r.$$

We address now the question.

Question 10.20.1
Does there exist a g such that

$$g^{q^r} \equiv 1 \pmod{p} \tag{10.62}$$

$$g^{q^{r-1}} \not\equiv 1 \pmod{p} \tag{10.63}$$

Equation (10.62) is equivalent to $g^{q^r} - 1 \equiv 0 \pmod{p}$. Thus g is a root of the $x^{q^r} - 1 \equiv 0 \pmod{p}$ and by the previous Corollary to Lagrange's theorem we know that there are q^r roots provided that $q^r|p-1$ which is indeed the case.
Equation (10.63) generates polynomial $x^{q^{r-1}} - 1 \equiv 0 \pmod{p}$. Another application of the previous Corollary to Lagrange's theorem indicates that there are q^{r-1} solutions to this modular equation since $q^{r-1}|q^r$ and $q^r|p-1$ imply $q^{r-1}|p-1$. Thus there exist $q^r - q^{r-1}$ of the former roots that satisfy the inequatlity of Eq. (10.63). Thus the way to find a g as needed is to go through the elements of $\mathbb{U}_p$ and answer positively the question posed. Note that $q^r - q^{r-1} > 1$ and thus a g can be found, eventually. ■

Exercise 958.

(a) Show that for a prime number p, $\mathbb{Z}_p$ has a primitive root mod p, or equivalently that $\mathbb{U}_p$ has a generator

$$\exists g \in \mathbb{U}_p : ord_p(g) = \phi(p). \tag{10.64}$$

(b) Show that $\mathbb{U}_p$ is cyclic.

Solution.
(a)
Case 1: $p = 2$**.** If $p = 2$, then 1 is a primitive root of $\mathbb{Z}_p = \{1\}$.
Case 2: $p > 2$ **is an odd prime.** Let $p-1$ have a prime factorization as given below.

$$p-1 = p_1^{a_1} \dots p_k^{a_k}$$

where $p_1 < \ldots < p_k$. Let us form

$$x^{p-1} - 1 = (x^{{p_1}^{a_1}} - 1) f(x)$$

The first factor has exactly $p_1^{a_1}$ distinct roots following the discussion on the Corollary of Lagrange's theorem. From the previous problem there exists a g_1 such that $ord_p(g_1) = p_1^{a_1}$. Repeating this argument for every $i = 2, 3 \dots, k$ and p_i we conclude that there exists g_i of order $p_i^{a_i}$. Then by Corollary 10.49 we have that $g_1 \cdot g_2 \cdot \ldots \cdot g_k$ has order $p_1^{a_1} \cdot \ldots p_k^{a_k} = p-1$, and thus $g_1 \dots g_k$ is a primitive root mod p.
(b) A primitive root for $\mathbb{U}_p$ implies an element of order $\phi(p)$. This mean $\mathbb{U}_p$ is cyclic. ■

Exercise 959.

Let p be a positive prime number. Suppose that $\phi(p) = p-1 = p_1^{a_1} \dots p_k^{a_k}$ are the prime factors of $p-1$, where $p_1 < \dots < p_k$ with $a_i > 0$.

Show that g is a primititve root $\pmod p$ if and only if $g^{\frac{p-1}{p_i}} \not\equiv 1 \pmod p$ for every p_i in other words,

$$g \text{ is primitive root mod } p \quad \Leftrightarrow \quad g^{\frac{p-1}{p_i}} \not\equiv 1 \pmod p, \forall p_i.$$

Example. If $p = 7$, then $p-1 = 6 = 2 \cdot 3$. To check that 2 is not a primitive root $\pmod 7$ we check whether $2^{(6/2)} \equiv 1 \pmod 7$. Given that $2^3 \equiv 1 \pmod 7$, this is confirmed. If 2 were a primitive root then $2^3 \not\equiv 1 \pmod 7$, an obvious fallacy.

Solution.

This is derived from a previous problem, and its Eq.(10.60) and the answer to its Question (10.20.1). ■

Exercise 960.

We extend the existence of primitive roots beyond prime numbers with the following question.
Primitive roots mod n and $2n$. Let n be equal to two ($n = 2$) or to an odd prime integer power $n = p^a$, $a \geq 1$. Then there exists a primitive root mod n if and only if there exists a primitive root mod $2n$.

Solution.
Let g be any odd integer. Then

$$g \equiv 1 \pmod{2}.$$

Then for any $k \geq 1$ we have the following.

$$g^k \equiv 1 \pmod{2}.$$

If $ord_n(g) = k$ then we have the following.

$$g^k \equiv 1 \pmod{n}.$$

Combining the two using Equation (10.45) we have the following.

$$g^k \equiv 1 \pmod{2n}.$$

Thus g is a primitive root mod n if and only if it is a primitive root mod $2n$. ■

Exercise 961.

Primitive roots mod p^a or $2p^a$. Let p be odd prime (positive) number. Let n be equal to two ($n=2$) or n is equal to an odd prime integer power $n=p^a$, $a\geq 1$. Then if there exists a primitive root mod n, then either $n=p^a$ or $n=2p^a$, with n as defined.

Solution.
Let p be an odd prime number. Let $n=Np^a$ for some integer $a\geq 1$. N is such that $\gcd(N,p^a)=\gcd(N,p)=1$ as otherwise we increase a and reduce N to N/p. Let $N\geq 3$. We will show that that there are no primitive roots mod n for $N\geq 3$. There will be then for $N=1$ and $N=2$. Consider the following.

$$\phi(n)=\phi(Np^a)=\phi(N)\phi(p^a).$$

Both ϕ are even numbers. For example $\phi(p^a)=p^{a-1}(p-1)$ and p is odd. Similar considerations apply to N. Let $b\in\mathbb{Z}$ and $\gcd(b,n)=1$ and let us assume it is a primitive root mod n. Moreover $\gcd(b,N)=1$ and then $b^{\phi(N)}\equiv 1 \pmod N$. Additionally $\gcd(b,p^a)=1$ and then $b^{\phi(p^a)}\equiv 1 \pmod p)^a$. We first obtain the following.

$$b^{\frac{\phi(n)}{2}}\equiv b^{\frac{\phi(N)\phi(p^a)}{2}}\equiv \left(b^{\phi(N)}\right)^{\frac{\phi(p^a)}{2}}\equiv 1 \pmod N.$$

We then obtain the following.

$$b^{\frac{\phi(n)}{2}}\equiv b^{\frac{\phi(N)\phi(p^a)}{2}}\equiv \left(b^{\phi(p^a)}\right)^{\frac{\phi(N)}{2}}\equiv 1 \pmod{p^a}.$$

Combining the two using Equation (10.45) we have the following.

$$b^{\frac{\phi(n)}{2}}\equiv 1 \pmod{Np^a}\Rightarrow b^{\frac{\phi(n)}{2}}\equiv 1 \pmod n.$$

But b is a primitive root mod n and thus the following applies.

$$a^{\phi(n)}\equiv 1 \pmod n$$

For a primitive root with order $\phi(n)$, the equality $a^{\frac{\phi(n)}{2}}\equiv 1 \pmod n$ contradicts it as it implies an order $\frac{\phi(n)}{2}<\phi(n)$. Thus there can be no primitive root for Np^a if $N\geq 3$. The only possible case(s) is/are p^a and $2p^a$. ■

Exercise 962.
(a) There are primitive roots mod n, where $n = 2$ or $n = 4$.

$$\exists g_2, \quad g_2 \text{ is a primitive root mod } \quad n = 2 \ \wedge \ \exists g_4, \quad g_4 \text{ is a primitive root mod } \quad n = 4 \ . \tag{10.65}$$

(b) There are no primitive roots mod n, where $n = 8$.

$$\nexists g_8, \quad g_8 \text{ is a primitive root mod } \quad n = 8. \tag{10.66}$$

Solution.
Direct inspection. For the former case $|U_2| = 1$ and contains 1. The primitive root is 1. For the latter case $|U_4| = 2$ and contains 1,3. The primitive root is 3.
(b) For $n = 8$, we have $|U_8| = 4$ and contains 1,3,5,7. For all these value $x = 1, 3, 5, 7$ we have $x^2 \equiv 1 \pmod 8$. For an x to be a primitive root we require $x^4 \equiv 1 \pmod 8$ i.e. the minimal power/order of x to be 4 and not 2, since $\phi(8) = 4$. ■

Exercise 963.

Let $n = 2^a$, where $a \geq 3$. Then, there are NO primitive roots mod n for $n = 2^a$ as defined.

$$\nexists g, \quad g \text{ is a primitive root mod } \quad n = 2^a,\ a \geq 3. \tag{10.67}$$

Solution.
We prove the result by induction on a. Base case $a = 3$. There are no primitive roots mod $n = 2^3 = 8$. This is part (b) of the previous problem and can use a proof by inspection.
For the inductive step we we then show by induction that

$$g^{2^{a-2}} \equiv 1 \pmod{2^a} \tag{10.68}$$

for all g such that $\gcd(g,n) = \gcd(g,2^a) = 1$. The latter implies an odd g.
The base case is $a = 3$ and Equation (10.68) becomes $g^2 \equiv 1 \pmod{2^3}$, which is true for every odd number g and thus every element of U_8.
For the inductive step, assuming $g^{2^{a-2}} \equiv 1 \pmod{2^a}$ we shall show the following.

$$g^{2^{a-1}} \equiv 1 \pmod{2^{a+1}}.$$

Equation (10.68) implies that g can never have order $\phi(2^a) = 2^{a-1}$, as $2^{a-2} < 2^{a-1}$. From Equation (10.68) we obtain that there exists an A such that

$$g^{2^{a-2}} - 1 = A2^a \Leftrightarrow g^{2^{a-2}} = 1 + A2^a \Leftrightarrow g^{2^{a-1}} = (1 + A2^a)^2 \Leftrightarrow g^{2^{a-1}} = 1 + A2^{a+1} + A^2A^{2a} \equiv 1 \pmod{2^{a+1}}.$$

This proves the inductive step, induction is incomplete and $g^{2^{a-2}} \equiv 1 \pmod{2^a}$ implies that any g of U_n, $n = 2^a$, $a \geq 3$ cannot be a primitive root mod n. ■

Exercise 964.

Primitive roots mod p^2, where p is an odd prime number. Show that if p is an odd prime number and thus $p > 2$, then $\mathbb{Z}_{p^2}$ has a primitive root.

$$\exists g, \quad g \text{ is a primitive root mod } \quad p^2,\ p > 2 \text{ is prime.} \tag{10.69}$$

Note. So far in prior problems we have proved the existence of primitive roots mod n, where n is an (odd) prime number, or $n = 2$ or $n = 4$. We also proved some auxiliar results such as there exists a primite root mod n if and only if there exists a primitive room mod $2n$, for $n = 2$ or $n = p^a$, p an odd prime number. We also proved that there are no primitive roots for $n = 2^a$, where $a \geq 3$.

Solution.
Let g be a primitive root mod p. Since g is a primitive root we have $g^{p-1} \equiv 1 \pmod{p}$. Let $k = ord_{p^2}(g)$. Since $g^k \equiv 1 \pmod{p^2}$ we also have $g^k \equiv 1 \pmod{p}$. Since g is a primitive root mod p, and $\phi(p) = p-1$ we also have that $p-1|k$. Thus $k = (p-1)r$. Moreover $k|\phi(p^2)$ since $k = ord_{p^2}(g)$ that is, $k|p(p-1)$. That is $p(p-1) = (p-1)rs$. This means $r|p$. Since p is prime this means $r = 1$ or $r = p$. (This by itself is also a previous problem.)
If $r = p$, $k = (p-1)r = p(p-1)$, then $\text{ord}_{p^2}(g) = k = p(p-1) = \phi(p^2)$. This means g **is a primitive root mod** p^2.
If $r = 1$, $k = (p-1)r = (p-1)$, then $ord_{p^2}(g) = k = p-1$ and this means $g^{p-1} \equiv 1 \pmod{p^2}$.
Consider $g_1 = g + p$. g_1 is also a primitive root $\pmod{p}$. This is because

$$g_1^{p-1} \equiv (g+p)^{p-1} \equiv g^{p-1} + \lambda p \equiv 1 + 0 \equiv 1 \pmod{p}$$

implies $g_1^{p-1} \equiv 1 \pmod{p}$. Similarly as before $\text{ord}_{p^2} g_1 = r_1(p-1)$ where r_1 is either 1 or p.
Consider that $g^{p-1} \equiv 1 \pmod{p^2}$ implies that

$$g_1^{p-1} \equiv (g+p)^{p-1} \equiv g^{p-1} + p(p-1)g^{p-2} + tp^2 g^{p-3} \equiv g^{p-1} - pg^{p-2} + (t + g^{p-2})p^2 \equiv 1 - pg^{p-2} \pmod{p^2}$$

If $r_1 = 1$ then $\text{ord}_{p^2}(g_1) = (p-1)r_1 = 1$, and thus

$$1 \equiv g_1^{p-1} \equiv 1 - pg^{p-2} \pmod{p}^2$$

which implies $pg^{p-2} \equiv 0 \pmod{p}^2$ i.e. $p|g$ i.e. $p \leq g$. This contradicts the fact that g chosen as primitive root mod p implies $g < p$. That is **it can't be that** $r_1 = 1$**.**
Thus $r_1 = p$ and $\text{ord}_{p^2}(g_1) = (p-1)r_1 = p(p-1) = \phi(p^2)$, and thus g_1 **is primitive root** $\pmod{p^2}$.
To conclude if $r = p$ then g is a primitive root $\pmod{p^2}$. If $r = 1$ then $g_1 = g + p$ implies that g_1 is a primitive root $\pmod{p^2}$. One way or the other there is a primitive root in $\mathbb{Z}_p^2$, and this is either a primite root g of $\mathbb{Z}_p$ or the sum of the primitive root plus p i.e. $g + p$!

In the next problem, we use this step as the base case or stepping stone of an inductive proof. ■

Exercise 965.

Let g be a primitive root mod p, where p is an odd prime. Let g be such that $g^{p-1} \not\equiv 1 \pmod{p^2}$. Show then the following.

$$g^{p-1} \not\equiv 1 \pmod{p^2} \Rightarrow g^{\phi(p^a)} \not\equiv 1 \pmod{p^{a+1}} \quad \forall a \geq 1. \tag{10.70}$$

We are going to show in the next few problems that there are primitive roots mod a prime power $n = p^a$, $a \geq 1$, or twice as much as that that is $n = 2p^a$. This step will assist in inductive steps of induction-based proofs.

Solution.
Proof by induction on a. Base case is $a = 1$ that is

$$g^{\phi(p)} \not\equiv 1 \pmod{p^2},$$

which is another writing of

$$g^{p-1} \not\equiv 1 \pmod{p^2},$$

since $\phi(p) = p - 1$.
Inductive step. We show the result for $a \geq 2$. If

$$g^{\phi(p^a)} \not\equiv 1 \pmod{p^{a+1}},$$

we show then

$$g^{\phi(p^{a+1})} \not\equiv 1 \pmod{p^{a+2}},$$

By Euler's theorem we have the following.

$$g^{\phi(p^a)} \equiv 1 \pmod{p^a} \Rightarrow g^{\phi(p^a)} - 1 = Ap^a \Rightarrow g^{\phi(p^a)} = 1 + Ap^a.$$

The induction hypothesis

$$g^{\phi(p^a)} - 1 \neq B \pmod{p^{a+1}},$$

implies that $p \nmid A$, since otherwise

$$g^{\phi(p^a)} - 1 = B \pmod{p^{a+1}}.$$

Consider

$$\phi(p^{a+1}) = p^a(p-1) = p \times \phi(p^a).$$

We then obtain

$$\begin{aligned} g^{\phi(p^{a+1})} &= g^{p\phi(p^a)} \\ &= (1+Ap^a)^p \\ &= 1 + Ap^{a+1} \\ &\not\equiv 1 \mod p^{a+2}, \end{aligned}$$

as needed to conclude the inductive step, where in the last derivation above one used the binomial theorem, for $x = Ap^a$, as shown below.

$$\begin{aligned} (1+Ap^a)^p = (1+x)^p &= \sum_{i=0}^{p} \binom{p}{i} x^i 1^{p-1} \\ &= 1 + px + x^2 \cdot S \\ &= 1 + Appp^a + p^{2a} \cdot S \\ &= 1 + Ap^{a+1} + p^{a+2} \cdot Sp^{a-2}. \end{aligned}$$

■

Exercise 966.

Primitive root mod p^a, $a \geq 2$. Show the following. Let g be a primitive root mod an odd positive prime number p. Then either g or $g+p$ is a primitive root mod p^a for all $a \geq 2$.
Note. We have already shown the result for $a = 2$.

Solution.
Let g be a primitive root mod p. Then we have the following.

$$g^{p-1} \equiv 1 \pmod{p},$$

with $\gcd(g, p) = 1$. We distinguish two cases.

$$\begin{array}{ll} \text{Case 1:} & g^{p-1} \not\equiv 1 \pmod{p^2} \\ \text{Case 2:} & g^{p-1} \equiv 1 \pmod{p^2} \Rightarrow \quad (p+g)^{p-1} \not\equiv 1 \pmod{p^2}. \end{array} \tag{10.71}$$

We will then show by induction that for all $a \geq 1$, we have the following.

$$ord_{p^a}(g) = \phi(p^a) = p^a - p^{a-1}, \tag{10.72}$$

that is, g is a primitive root mod p^a.
Case 1, base case $a = 1$. Equation (10.72) is true for $a = 1$ as a by product of Fermat's little theorem which also establishes that g is a primitive root mod p. This is the basis of the inductive proof.
Case 1, inductive step from a to $a+1$. We assume that Equation (10.72) is true for a, we shall show it is true for $a+1$. That is, the following will be shown.

$$ord_p(g) = \phi(p) \quad \wedge \quad ord_{p^a}(g) = \phi(p^a) \Rightarrow ord_{p^{a+1}}(g) = \phi(p^{a+1}).$$

We will show that

$$k = ord_{p^{a+1}}(g) \quad \wedge \quad g^k \equiv 1 \pmod{p^{a+1}} \Rightarrow k = \phi(p^{a+1}).$$

Since $g^k \equiv 1 \pmod{p^{a+1}}$ this implies the following

$$g^k - 1 = Kp^{a+1} \Rightarrow p^a | g^k - 1 \Rightarrow g^k \equiv 1 \pmod{p^a} \Rightarrow ord_{p^a}(g) | k \Rightarrow \phi(p^a) | k$$

Furthermore,

$$g^k \equiv 1 \pmod{p^{a+1}} \Leftrightarrow ord_{p^{a+1}}(g) | \phi(p^{a+1}) \Leftrightarrow k | \phi(p^{a+1})$$

From the two derivations $\phi(p^a)|k$ and $k|\phi(p^{a+1})$ we conclude that (a) either $k = \phi(p^{a+1}) = p^a(p-1)$ (b) or $k = \phi(p^a)) = p^{a-1}(p-1)$.
Case (b) leads to contradiction. If it is not the former case (a), then it is the latter case (b) and $k = \phi(p^a)) = p^{a-1}(p-1)$. The latter implies

$$k = ord_{p^{a+1}}(g) = \phi(p^a)) \quad \wedge \quad g^{\phi(p^a)} \equiv 1 \pmod{p^{a+1}},$$

which contradicts Eq.(10.70) that from

$$g^{p-1} \not\equiv 1 \pmod{p^2},$$

derives a

$$g^{\phi(p^a)} \not\equiv 1 \pmod{p^{a+1}}$$

instead. Thus case (a) is applicable and we conclude that g has order $k = \phi(p^{a+1})$ thus completing the inductive step. As a conclusion g is a primitive root mod p^a for all $a \geq 1$.
Case 2, base case $a = 1$. Equation (10.72) is true for $a = 1$ as a by product of Fermat's little theorem which also establishes that g is a primitive root mod p. This is the basis of the inductive proof. Furthermore, by way of Case 2 we have the following.

$$g^{p-1} \equiv 1 \pmod{p^2}.$$

Case 2, inductive step from a to $a+1$. We then consider $g+p$ as a candidate for a primitive root. We note that since g is a primitive root mod p so is $p+g$. We use the binomial theorem for

$$\begin{aligned}(p+g)^{p-1} &= \sum_{i=0}^{p-1}\binom{p-1}{i}p^i g^{p-1-i}\\ &= p^0 g^{p-1}+(p-1)pg^{p-2}+p^2P\\ &\equiv g^{p-1}+p^2g^{p-2}-pg^{p-2} \pmod{p^2}\\ &\equiv g^{p-1}-pg^{p-2} \pmod{p^2}\\ &\equiv 1-pg^{p-2} \pmod{p^2}.\end{aligned}$$

where we used the $g^{p-1}\equiv 1 \pmod{p^2}$ to obtain the last derivation above. Moreover $\gcd(p,g)=1$ and further more we obtain

$$(p+g)^{p-1}\equiv 1-pg^{p-2} \pmod{p^2} \not\equiv 1 \pmod{p^2}.$$

This is a condition symmetric to the one of Case 1. Using similar arguments we conclude that $p+g$ is primitive root for this case mod p^a, for every $a\geq 1$. ■

Exercise 967.

Primitive roots mod p^a, $a \geq 1$. Show that there are primitive roots mod p^a for all $a \geq 1$, where p is an odd prime number.

$$\exists g, \quad g \text{ is a primitive root mod } \quad p^a, \ p > 2 \text{ is prime, } a \geq 1, a \in \mathbb{Z}. \tag{10.73}$$

Solution.
From the previous problem if g is a primitive root mod p, then either g or $g+p$ is a primitive root mod p^a for all $a \geq 1$. By Equation (10.64) we know that there exists a primitive root g mod p. Thus there exist primitive roots mod p^a. Furthermore if there exists one primitive root mod p, from Eq. (10.52) there are $\phi(\phi(p)) = \phi(p-1)$ such primitive roots mod p and by extension mod p^a. Every primitive root mod p is a generator of $\mathbb{U}_p$ and thus $\mathbb{U}_p$ is cyclic. ■

Exercise 968.

Primitive roots mod $2p^a$, $a \geq 1$. Show that there are primitive roots mod $2p^a$ for all $a \geq 1$.

$$\exists g, \quad g \text{ is a primitive root mod } \quad 2p^a, \ p > 2 \text{ is prime, } a \geq 1, a \in \mathbb{Z}. \tag{10.74}$$

This is an extension of the previous problem.

Solution.
This is going to be proved indirectly. By the previous problem we know that there is a primitive root, call it g, mod p^a, $a \geq 1$.
We are going to show that if g is odd then g is also a primitive root mod $2p^a$. Otherwise, $g + p^a$ is odd, and in this case we can show that $g + p^a$ is a primitive root mod $2p^a$.
Case 1: g is odd. Then, the sequence,

$$g, g^2, g^3, \dots g^{\phi(p^a)},$$

is a sequence of odd integer and are distinct mod p^a. Then they are also distinct mod $2p^a$ as well. Note that

$$\phi(2p^a) = \phi(2)\phi(p^a) = \phi(p^2).$$

Thus the sequence above enumerates all unit not only of $\mathbb{U}_{p^a}$ but also $\mathbb{U}_{2p^a}$ since

$$|\mathbb{U}_{p^a}| = |\mathbb{U}_{2p^a}| = \phi(p^a).$$

Case 2 g is even. The $g + p^a$ is odd and proceed likewise.
This concludes the problem. ■

The conclusion leads to the following theorem.

Theorem 10.14

There are primitive roots mod n where $n = 2, 4, p^a, 2p^a$ for $a \geq 1$. The number of primitive roots mod n is $\phi(\phi(n))$.

The latter can be show independently of the results for $n = 2, 4, p^a, 2p^a$ for $a \geq 1$. See the following problem.

Exercise 969.

Let $n \in \mathbb{Z}_+$ and let us assume that there exists a primitive root mod n. Then $\phi(\phi(n))$ of the numbers of $\mathbb{U}_p$ that is $1, 2, \ldots, n-1$ are primitive roots mod n.

Solution.

If g is a primitive root mod n, then g becomes a generator of $\mathbb{U}_p$. We need to compute k such that for $1 \leq k \leq \phi(n)$, $a^k \pmod{n}$ is a generator for $\mathbb{U}_p$. a^k is a generator if and only if $a^{kl} \equiv a^m \pmod{n}$, for every $1 \leq m \leq \phi(n)$, has a solution for every l, as needed. For this to happen $kl \equiv m \pmod{\phi(n)}$ must have a solution. For the latter we need $\gcd(k, \phi(n)) | m$ for all values of m, that is for $\gcd(k, \phi(n)) = 1$. Therefore the cardinality of the set

$$\{1 \leq k \leq \phi(n) : \gcd(k, \phi(n)) = 1\}$$

is $\phi(\phi(n))$, as needed. ■

Exercise 970.

Alternative: primitive roots mod $2,4,p^a,2p^a$. Show that $\mathbb{Z}_n$ has a primitive root if and only if $n=2,4,p^a,2p^a$ where p is an odd prime number and a is a positive integer.

Solution.

Case 1: $n=2,4$. By way of Eq.(10.65 there exist primitive roots mod $n=2$ and mod $n=4$. They are by direct inspection 1 and 3 respectively.

Case 2: $n=p^a$, $a\geq 1$. This is a by product of Eq.(10.73).

Case 3: $n=2p^a$, $a\geq 1$. This is a by product of Eq.(10.74). ■

Exercise 971.
Find all the primitive roots mod 11.

Solution.
Let $n = 11$. Clearly $\phi(11) = 10$ and $\phi(\phi(11)) = \phi(10) = \phi(2)\phi(5) = 1 \cdot 4 = 4$. If 2 is a generator by a prior problem 2^k is a generator for those k such that $\gcd(k, \phi(n)) = 1$. For the latter, and $n = 11$, we have $\gcd(k, \phi(11)) = \gcd(k, 10) = 1$ for $k = 1, 3, 7, 9$. Thus $2^1, 2^3, 2^7 \pmod{11}, 2^9 \pmod{11}$ all four generators and thus primitive roots, given that $\phi(\phi(11)) = 4$. ■

10.21 Legendre and Jacobi symbols

10.21.1 Summary of Legendre symbol properties

Note 10.21.1

We write q.r. to indicate that an a is a quadratic residue and nq.r. to indicate that it is not a quadratic residue i.e. it is a quadratic non-residue.

Note 10.21.2

The set of units $(\mathbb{Z}/n\mathbb{Z})^x$ for prime n will be represented by $\mathbb{U}_n$. Then U_n becomes a field. For $g \in \mathbb{U}_n$ it is $\gcd(g,n) = 1$ and g is invertible and thus $gx \equiv 1 \pmod{n}$ exists, and it is known as a unit. Moreover $\mathbb{U}_n$ for prime n it is a cyclic group.

The Legendre symbol $\left(\frac{a}{p}\right)$ is defined to be 1 if a is a quadratic residue $\pmod{p}$. It is -1 if a is a quadratic non-residue (i.e. it is not a quadratic residue mod p). Therefore for those two cases $p \nmid a$. Furthermore, it is 0 if $p|a$.

Definition 10.59 **Legendre symbol**

Let $p \in \mathbb{N}$ be an odd prime number. For any $a \in \mathbb{Z}$ the Legendre symbol is defined as follows.

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if} p|a \\ 1 & \text{if a is quadratic residue mod p} \\ -1 & \text{otherwise.} \end{cases}$$

Proposition 10.13 **Legendre symbol properties**

Let $p \in \mathbb{N}$ be an odd prime number ($p > 2$). For any $a, b \in \mathbb{Z}$ the following apply.

$$a \equiv b \pmod{p} \Rightarrow \left(\frac{a}{p}\right) = \left(\frac{b}{p}\right). \tag{10.75}$$

$$\gcd(a,p) = 1 \Rightarrow \left(\frac{a^2}{p}\right) = 1. \tag{10.76}$$

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{10.77}$$

$$\gcd(a,p) = 1, a \text{ is q.r.} \Leftrightarrow a^{\frac{p-1}{2}} \equiv 1 \pmod{p} \tag{10.78}$$

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}. \tag{10.79}$$

$$\left(\frac{-1}{p}\right) = 1 \Leftrightarrow p \equiv 1 \pmod{4}. \tag{10.80}$$

$$\left(\frac{-1}{p}\right) = -1 \Leftrightarrow p \equiv 3 \pmod{4}. \tag{10.81}$$

$$\sum_{a=0}^{p-1} \left(\frac{a}{p}\right) = 0. \tag{10.82}$$

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = 1 \Rightarrow \left(\frac{ab}{p}\right) = 1. \tag{10.83}$$

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = -1 \Rightarrow \left(\frac{ab}{p}\right) = 1. \tag{10.84}$$

(10.85)

Proposition 10.14 **Legendre symbol properties**

Let $p,q \in \mathbb{N}$ be odd prime numbers $(p,q>2)$. The following apply.

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}. \tag{10.86}$$

$$\left(\frac{p}{q}\right) = \begin{cases} -\left(\frac{q}{p}\right) & p \equiv q \equiv 3 \pmod 4 \\ \left(\frac{q}{p}\right) & \text{otherwise} \end{cases} \tag{10.87}$$

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}. \tag{10.88}$$

$$\left(\frac{2}{p}\right) = 1 \Leftarrow p \equiv \pm 1 \pmod 8. \tag{10.89}$$

$$\left(\frac{2}{p}\right) = -1 \Leftarrow p \equiv \pm 3 \pmod 8. \tag{10.90}$$

(10.91)

Proposition 10.15 **Legendre symbol properties**

Let $p \in \mathbb{N}$ be an odd prime number $(p>2)$. Let $a \in \mathbb{Z}_p^*$ such that $a>0$ and $\gcd(a,p)=1$. Consider

$$I(a) = \{1 \le i \le (p-1)/2 : ia \pmod p\}. = \{1 \le i \le (p-1)/2 : R(i)\},$$

and

$$T = \{1 \le i \le (p-1)/2 : R(i) < 0\},$$

where $R(i)$ is the absolute least residue of $ai \pmod p$, and thus $-(p-1)/2 \le r(i) \le (p-1)/2$, with

$$R(i) \equiv i \cdot a \equiv a(i)r(i) \pmod p,$$

and $a(i) \equiv \pm 1 \pmod p$.

$$\left(\frac{a}{p}\right) = (-1)^{|T|} \qquad \text{(Gauss)}. \tag{10.92}$$

$$a \ne p, a \bmod 2 = 1 : \left(\frac{a}{p}\right) = (-1)^{\sum_{i=1}^{\frac{p-1}{2}} \lfloor \frac{ia}{p} \rfloor} \qquad \text{(Eisenstein)}. \tag{10.93}$$

10.21.2 Summary of Jacobi symbol properties

Note 10.21.3

In all cases below, n is odd and positive.

Definition 10.60 **Jacobi symbol**

Let $n \in \mathbb{N}$ and $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$. Let n be a product of odd prime numbers, not necessarily distinct,

$$n = p_1 p_1 \dots p_k,$$

then the Jacobi symbol is defined as follows.

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \cdot \left(\frac{a}{p_2}\right) \cdot \ldots \cdot \left(\frac{a}{p_k}\right) \cdot .$$

The Jacobi symbol for a prime number n is the Legendre symbol.

Note 10.21.4

Note that for the Jacobi symbol, we have $\gcd(a,n) = 1$. If a is such that $\gcd(a,n) \neq 1$, because of the prime decomposition of n we would have $p_i | a$ for some i. Then $\left(\frac{a}{p_i}\right) = 0$ and consequently $\left(\frac{a}{n}\right) = 0$.

Proposition 10.16 **Jacobi symbol properties**

Let $n, m \in \mathbb{N}$, $n > 0, m > 0$ be odd integers. Let $a \in \mathbb{Z}$ such that $\gcd(a,n) = 1$. Furthermore, let $a, b \in \mathbb{Z}$ such that $\gcd(nm, ab) = 1$. The following then apply.

$$a \text{ is a q.r. mod n} \Rightarrow \left(\frac{a}{n}\right) = 1. \tag{10.94}$$

$$a \equiv b \pmod{n} \Rightarrow \left(\frac{a}{n}\right) = \left(\frac{b}{n}\right). \tag{10.95}$$

$$\left(\frac{a}{n}\right)\left(\frac{a}{m}\right) = \left(\frac{a}{nm}\right). \tag{10.96}$$

$$\left(\frac{a}{n}\right)\left(\frac{b}{n}\right) = \left(\frac{ab}{n}\right). \tag{10.97}$$

$$\left(\frac{a^2}{n}\right) = \left(\frac{a}{n^2}\right) = 1. \tag{10.98}$$

$$\left(\frac{a^2 b}{n^2 m}\right) = \left(\frac{b}{m}\right). \tag{10.99}$$

$$\left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}. \tag{10.100}$$

$$\left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}. \tag{10.101}$$

$$\gcd(n,m) = 1, \left(\frac{n}{m}\right)\left(\frac{m}{n}\right) = (-1)^{\frac{n-1}{2}\frac{m-1}{2}}. \tag{10.102}$$

Let m, m be odd integer numbers such that $\gcd(n,m) = 1$, where $n = n_1 n_2 \dots n_k$, $m = m_1 m_2 \dots m_l$, for $k \geq i \geq 1$, $l \geq j \geq 1$ and n_i, m_j prime numbers.

$$\left(\frac{n}{m}\right) = \prod_i \prod_j \left(\frac{n_i}{m_j}\right). \tag{10.103}$$

Exercise 972.

If p,q are odd integers, then the following applies.
(a)

$$\frac{p-1}{2}+\frac{q-1}{2}\equiv\frac{pq-1}{2}\mod 2. \tag{10.104}$$

This can be generalized for $p_1,p_2,\dots p_k$, with $p=p_1\cdot p_2\cdot\ldots\cdot p_k$.
(b)

$$\frac{p_1-1}{2}+\frac{p_2-1}{2}+\ldots+\frac{p_k-1}{2}\equiv\frac{p-1}{2}\mod 2. \tag{10.105}$$

Solution.
(a) Let $p=2P+1$, $q=2Q+1$ for odd p,q. Then

$$\frac{pq-1}{2}=\frac{(2P+1)(2Q+1)-1}{2}=2PQ+P+Q\equiv 0+P+Q\pmod 2\equiv 0+\frac{p-1}{2}+\frac{q-1}{2}\pmod 2.$$

(b) Follows by induction from part (a). ■

Exercise 973.

If p,q are odd integers, then the following applies.
(a)

$$\frac{p^2-1}{8}+\frac{q^2-1}{8}\equiv\frac{p^2q^2-1}{8}\mod 2. \tag{10.106}$$

This can be generalized for $p_1,p_2,\dots p_k$, with $p=p_1\cdot p_2\cdot\ldots\cdot p_k$.
(b)

$$\frac{p_1^2-1}{8}+\frac{p_2^2-1}{8}+\ldots+\frac{p_k^2-1}{8}\equiv\frac{p^2-1}{8}\mod 2. \tag{10.107}$$

Solution.
(a) Let $p=2P+1$, $q=2Q+1$ for odd p,q. Then

$$p^2-1=4P(P+1),\quad q^2-1=4Q(Q+1)\Rightarrow\frac{p^2-1}{8}=\frac{P(P+1)}{2},\quad\frac{q^2-1}{8}=\frac{Q(Q+1)}{2},$$

and the products $P(P+1)$, $Q(Q+1)$ are even integers. Moreover

$$\frac{p^2-1}{8}+\frac{q^2-1}{8}=\frac{P(P+1)}{2}+\frac{Q(Q+1)}{2}.$$

Similarly

$$p^2q^2=16P^2Q^2+16P^2Q+4P^2+16PQ^2+16PQ+4P+4Q^2+4Q+1,$$

and therefore

$$\frac{p^2q^2-1}{8}=(2P^2Q^2+2P^2Q+2PQ^2+2PQ)+\frac{P(P+1)}{2}+\frac{Q(Q+1)}{2}.$$

which, given that the parenthesized term is an even number, leads to the following.

$$\frac{p^2q^2-1}{8}\equiv 0+\frac{P(P+1)}{2}+\frac{Q(Q+1)}{2}\pmod 2,$$

$$\frac{p^2q^2-1}{8}\equiv 0+\frac{P(P+1)}{2}+\frac{Q(Q+1)}{2}=\frac{p^2-1}{8}+\frac{q^2-1}{8}\pmod 2.$$

(b) Follows by induction from part (a). ■

10.21.3 Legendre symbol properties

Exercise 974.

Legendre symbol properties. Show that if $a \equiv b \pmod{p}$ for an odd prime number p, then

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$$

Note. In the remainder q.r. mean quadratic residue, and q.nr. means quadratic non-residue.

Solution.
Case 1: $p|a$**.** Note that if $p|a$ then since $a \equiv b \pmod{p}$ we have $p|b$ and thus in that case

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right) = 0$$

Case 2: a is a q.r.
Let a otherwise be a q.r. mod p i.e.

$$\left(\frac{a}{p}\right) = 1.$$

Then there exists an x such that $x^2 \equiv a \pmod{p}$ and thus $p|x^2 - a$. Moreover $a \equiv b \pmod{p}$ implies $p|a-b$ and by addition $p|x^2 - b$ i.e. $x^2 \equiv b \pmod{p}$ i.e. b is a q.r. mod p. Same if b is a q.r. mod p.

Case 3: a is a q.nr.
Let a be a q.nr. mod p. Then there is NO x such that $x^2 \equiv a \pmod{p}$. This implies that b is also a q. nr. mod p. Why ? Otherwise if b is a q.r. mod p there exists a y such that $y^2 \equiv b \pmod{p}$ and by using $p|a-b$ we would then conclude a is a q.r. mod p, a contradiction. Case completed, cases completed, problem completed. ■

Exercise 975.

Legendre symbol more properties. Let p be an odd prime number > 2. Then for every $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$ then show the following.

$$\left(\frac{a^2}{p}\right) = 1.$$

Solution.
Integer a^2 is a quadratic residue mod p since there exists an x such that $x^2 \equiv a^2 \mod p$. This x is $x = a$ or $a \mod p$ in general if $a \geq p$. ■

Exercise 976.

More properties of the Legendre symbol. Let p be an odd prime number > 2. For all $a, b \in \mathbb{Z}_p$ we have

(a) If $a \equiv b \pmod{p}$ then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(b) $\left(\frac{a^2}{p}\right) = 1$.

(c) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.

Solution.
(a) It is immediate (and proven earlier).
(b) $(a^2)^{(p-1)/2} \equiv 1 \pmod{p}$ by Fermat's Little Theorem and Euler's criterion. Moreover, a^2 is such that $\left(\frac{a^2}{p}\right) = 1$ obviously by way of part (c).
(c) $\left(\frac{ab}{p}\right) \equiv (ab)^{(p-1)/2} \equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \pmod{p}$. Integer p is a prime number; furthermore the Legendre symbols on the left and right side of the equivalence are of a diference between -2 and 2. Give that p divides that difference the only possibility is for them to be equal to each other and thus

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

The result is true for all $a, b > 0$ as long as $\gcd(a, p) = \gcd(b, p) = 1$.

■

Exercise 977.
Let p be an odd prime. Show the following: If a is a quadratic residue mod p and b is a quadratic residue mod p then ab is a quadratic residue mod p as well.

Solution.
If a is a q.r. mod p then there exists x such that $x^2 \equiv a \pmod{p}$. If b is a q.r. mod p then there exists y such that $y^2 \equiv b \pmod{p}$. Then obviously $x^2y^2 = (xy)^2 \equiv ab \pmod{p}$, implying that ab is a quadratic residue mod p as well. ■

Exercise 978.

Let p be an odd prime number. The congruence relation

$$x^2 \equiv a \pmod{p} \tag{10.108}$$

has the following properties.
(1) If $a = 0$ then it has only one solution $x = 0 \bmod p$.
(2) If $p \nmid a$, then it has zero (0) or two (2) solutions mod p.

Solution.
Case (1). If $a = 0$, which implies $p|a$, then the congruence relation becomes as follows.

$$x^2 \equiv 0 \pmod{p},$$

which implies $p|x^2$ or equivalently $p|x$ and thus $x \equiv 0 \pmod{p}$, and the result follows.
Case (2). Suppose $p \nmid a$. If the modular equation has 0 solutions we are done. Otherwise it has at least one solution. One can use Lagrange's theorem to conclude that modular Eq.(10.108) for prime p has at most two distinct solutions. Consider x the "at least one solution" such that

$$x^2 \equiv a \pmod{p}$$

We observe then that $p - x$ is another solutions as follows.

$$(p-x)^2 \equiv p^2 - 2xp + x^2 \equiv 0 + x^2 \equiv x^2 \pmod{p}.$$

Thus if $x^2 \equiv a \pmod{p}$ then $(p-x)^2 \equiv a \pmod{p}$. Therefore we found a second solution mod p. It then suffices to show that those two solutions are distinct mod p to tie with the Lagrange upper-bound of two solutions. In order to prove $x \not\equiv p - x \pmod{p}$ let us assume that it is $x \equiv p - x \pmod{p}$. Then equivalently

$$x \equiv p - x \pmod{p} \Leftrightarrow 2x \equiv 0 \pmod{p} \Leftrightarrow p|x \Leftrightarrow p|x^2 \Leftrightarrow p|a,$$

since p is an odd prime and thus it can't divide two or be divisible by two. The conclusion $p|a$ contradicts $p \nmid a$. Thus x and $p - x$ are distinct mod p, if one of them exists. There are no more by Lagrange's theorem.
(We can sideswipe references to Lagrange's theorem by using an alternative arguement. Consider another solutions z to $z^2 \equiv a \pmod{p}$. Then $z^2 \equiv x^2 \equiv a \pmod{p}$ would lead to $(z-x)(z+x) \equiv 0 \pmod{p}$ and thus any other solution is either $z \equiv x \pmod{p}$ or $z \equiv -x \pmod{p}$ i.e. already included in the (set of) two solutions already obtained: x and $p - x$ mod p. We also show that if $x + y \neq p$ or $x + y \not\equiv 0 \pmod{p}$ then $x^2 \not\equiv y^2 \pmod{p}$ and thus the two solutions x and $p - x$ is all that we can have. If $x^2 \equiv y^2 \mod p$ we have as before $x - y \equiv 0 \pmod{p}$ or $x + y \equiv 0 \pmod{p}$. The latter contradicts $x + y \not\equiv 0 \pmod{p}$, and is dismissed. The former leads to $p|x - y$. But $-p < x - y < p$ and p primes leads to $p \nmid x - y$ which contradicts the former $x - y \equiv 0 \pmod{p}$, and is also dismissed. Thus we have covered all possible solutions.)
Furthermore, consider the set

$$\{1, 2, \ldots, i, \ldots, \frac{p-1}{2}, \frac{p+1}{2}, \ldots, p-1\}$$

and then, the set of

$$1^2, 2^2, \ldots, i^2 \pmod{p}, \ldots, (\frac{(p-1)}{2})^2 \pmod{p}$$

describes all $\frac{p-1}{2}$ quadratic residues mod p, with the remainder $\frac{p-1}{2}$ elements of the set forming the quadratic non-residues mod p. ■

10.21.4 Euler's criterion

Exercise 979.

Euler's criterion: simplified form. Let p be an odd prime number, for every $a \in \mathbb{Z}$ with $\gcd(a,p)=1$,

$$a^{\frac{p-1}{2}} = \begin{cases} 1 & \text{if a is quadratic residue mod p} \\ -1 & \text{if a is not a quadratic residue mod p.} \end{cases}$$

Solution.
Let $b = a^{\frac{p-1}{2}}$. Then

$$b^2 = a^{p-1} \equiv 1 \pmod{p},$$

by Fermat's Little theorem. The polynomial on (with indeterminate) b, by Lagrange's theorem has at most two roots. We recognize them as $1, p-1$ mod p i.e. $1, -1$ mod p.
Case 1. Let a be a q.r. mod p. Then

$$\exists x : x^2 \equiv a \pmod{p}.$$

Moreover for prime p, $\gcd(x,p)=1$ since $1 \leq x < p$, and by Fermat's Little theorem we have $x^{p-1} \equiv 1 \pmod{p}$. Then

$$a^{\frac{p-1}{2}} = x^{2\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p},$$

We proved the top part that $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ for a quadratic residue a mod p.
Case 2. Let a not be a q.r. mod p. Then set $z = a^{\frac{p-1}{2}}$. We know by the discussion prior to case 1, (rename x there to z or the other way around) that z is either $+1$ or -1 mod p. The equation $y^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ has at most $(p-1)/2$ solutions, after noting $(p-1)/2$ is an integer for odd p. Each such solution is a q.r. mod p by way of case 1. Thus the only possibility for a is the -1 case and thus $z = a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$. We thus proved the bottom part that $a^{\frac{p-1}{2}} \equiv -11 \pmod{p}$ for a non quadratic residue a mod p. ■

Exercise 980.

Euler's criterion: simpler form. Let p be an odd prime number, for every $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$, a is a quadratic residue mod p if and only if

$$\text{a is q.r. mod } p \Leftrightarrow a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Solution.
$\Rightarrow$. Let a be a q.r. mod p. Then

$$\exists x : x^2 \equiv a \pmod{p}.$$

Moreover for prime p, $\gcd(x,p) = 1$ since $1 \leq x < p$, and by Fermat's Little theorem we have $x^{p-1} \equiv 1 \pmod{p}$. Then

$$a^{\frac{p-1}{2}} = x^{2\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p},$$

and the only-if has been proved.
$\Leftarrow$. Let $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. We shall show that a is q.r. mod p.
If g is a primite root mod p, the $ord_p(g) = p-1$ and therefore

$$\exists k : g^k \equiv a \pmod{p} \Rightarrow g^{k(p-1)/2} \equiv a^{(p-1)/2} \pmod{p} \Rightarrow g^{k(p-1)/2} \equiv 1 \pmod{p}.$$

The latter implies $k(p-1)/2 \equiv 0 \pmod{p-1}$ or in other words $k = 2l$ i.e. k is even and thus $k/2$ an integer. Then

$$(g^{k/2})^2 \equiv a \pmod{p},$$

implying that a is a quadratic residue and the x of $x^2 \equiv a \pmod{p}$ is $x = g^{k/2} \pmod{p}$. ■

Exercise 981.

Simplified Euler's criterion for qudratic residuosity mod p.

Let p be an odd prime number, for every $a \in \mathbb{Z}$ the (simplified) Legendre symbol is defined as follows.

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if a is quadratic residue mod p} \\ -1 & \text{otherwise.} \end{cases}$$

Show then the following.

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}. \tag{10.109}$$

Solution.

We skip the possibility $p|a$ for which $\left(\frac{a}{p}\right) = 0$ by definition. For odd prime p then $(p-1)/2$ is an integer, and thus $p|a$ implies $p|a^{(p-1)/2}$ from which $a^{(p-1)/2} \equiv 0 \pmod{p}$ and thus $a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) \pmod{p}$.

Consider $x = a^{(p-1)/2}$, for some a such that $1 \leq a < p$. By Fermat's Little theorem $x^2 = a^{p-1} \equiv 1 \pmod{p}$. By Lagrange's theorem $x^2 \equiv 1 \pmod{p}$, for prime p has at most two solutions mod p. We can verify $+1$ and $-1 \equiv p-1 \pmod{p}$ as those two mod p.

We show that if a is a quadratic residue mod p, then

$$a^{(p-1)/2} \equiv 1 \pmod{p}.$$

If a is a quadratic residue, there exists b such that $1 \leq b < p$ such that $b^2 \equiv a \pmod{p}$. Then

$$a^{(p-1)/2} = (b^2)^{(p-1)/2} = b^{p-1} \equiv 1 \pmod{p},$$

by Fermat's Little theorem.

Let now a is NOT a quadratic residue mod p. The

$$x^{(p-1)/2} \equiv 1 \pmod{p},$$

has at most $(p-1)/2$ solutions. All these are the quadratic residues mod p. Note that for odd prime p then $(p-1)/2$ is an integer. Since a now is not a quadratic residue, it is not one of them, and thus

$$a^{(p-1)/2} \not\equiv 1 \pmod{p}.$$

Consider

$$y = a^{(p-1)/2},$$

As derived previously $y^2 \equiv 1 \pmod{p}$, by Fermat's Little theorem, and the only possibilities of $y^2 \equiv 1 \pmod{p}$ by Lagrange's theorem are $y \equiv 1$ or $y \equiv -1$ mod p. The latter is not possible for a quadratic non-residue, thus the only possibility left is $y \equiv -1$ mod p i.e.

$$y = a^{(p-1)/2} \equiv -1 \pmod{p}.$$

■

Exercise 982.

Euler's criterion for qudratic residuosity mod p.
More Legendre symbol properties. Show then the following.

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}. \tag{10.110}$$

Solution.
Case 1: $p|a$. Since p is an odd prime number, $(p-1)/2$ is an integer. Then $p|a$ implies $p|a^{\frac{p-1}{2}}$. The latter implies

$$a^{\frac{p-1}{2}} \equiv 0 \pmod{p},$$

and thus for $p|a$ it is indeed

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p} \Rightarrow \left(\frac{a}{p}\right) = 0.$$

Case 2: $p \nmid a$. In the remaining cases $p \nmid a$. Then $\gcd(a,p) = 1$ and by Fermat's little theorem we have

$$a^{p-1} \equiv 1 \pmod{p}.$$

By factoring we obtain the following

$$(a^{\frac{p-1}{2}} - 1)(a^{\frac{p-1}{2}} + 1) \equiv 0 \pmod{p}$$

The (congruence) equation

$$x^2 \equiv 1 \pmod{p}$$

has at most two distinct solutions by Lagrange's theorem mod p: $\pm 1 \pmod{p}$ and there are no more mod p. Furthermore

$$(a^{\frac{p-1}{2}} - 1)(a^{\frac{p-1}{2}} + 1) \equiv 0 \pmod{p} \Rightarrow \quad a^{\frac{p-1}{2}} \equiv 1 \pmod{p} \vee a^{\frac{p-1}{2}} \equiv -1 \pmod{p} \Rightarrow \quad \left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

We need to distinguish which case maps to residuosity and which one to non-residuosity.
Case 2a: a is a quadratic residue mod p, mapping to $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. Then there exists an $0 < x < p$ such that

$$x^2 \equiv a \pmod{p}.$$

We then have

$$a^{\frac{p-1}{2}} \equiv (x^2)^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p}.$$

This follows from Fermat's little Theorem as $0 < x < p$, p prime and thus $\gcd(x,p) = 1$. (Note also that if $p|x$ then $p|x^2$ and since $p|x^2 - a$ then $p|a$ contradicting $p \nmid a$.)
Moreover $z^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ has at most $\frac{p-1}{2}$ distinct solutions mod p, a is one of them and there are no more than $(p-1)/2$, and by the previous problem we know there are exactly $(p-1)/2$ distinct solutions, the quadratic residues mod p, which are $1^2, 2^2, \ldots, ((p-1)/2)^2 \bmod p$.
Case 2b: a is a quadratic non residue mod p, mapping to $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$. Let a be a quadratic non-residue. Since the $(p-1)/2$ quadratic residues satisfy $z^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ then since it is still $a^{p-1} \equiv 1 \pmod{p}$ this can only be satisfied by way of $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$ and thus the remaining $(p-1)/2$ quadratic non-residues will satisfy $z^{\frac{p-1}{2}} \equiv -1 \pmod{p}$ and this concludes the problem.
One can also argue as follows. For every $b = 1, 2, \ldots, p-1$ the congruence $bx \equiv a \pmod{p}$ is such that $\gcd(b,p) = 1$ and thus the congruence has a unique solution in $1, 2, \ldots, p-1$. Since a is a quadratic non residue we can't have $x = b$ since then $b \cdot b \equiv a \pmod{p}$ that would make a a quadratic residue. Thus the integers $1, 2, \ldots, p-1$ can be broken into pairs whose products are all equal to a. There are $(p-1)/2$ such pairs. Then by multiplying

all of them we get $(p-1)! \equiv a^{(p-1)/2} \pmod{p}$. The latter $(p-1)!$ is -1 by Wilson't theorem. Therefore for a quadratic non residue a we have the following

$$a^{(p-1)/2} \equiv -1 \pmod{p}.$$

■

10.21.5 More Legendre

Exercise 983.

More Legendre symbol properties: applications of Euler's criterion. Let p be prime number number > 2. Show that -1 is a quadratic residue $\pmod p$ if and only if $p \equiv 1 \pmod 4$.

$$\left(\frac{-1}{p}\right) = 1 \Leftrightarrow p \equiv 1 \pmod 4$$

Moreover

$$\left(\frac{-1}{p}\right) = -1 \Leftrightarrow p \equiv 3 \pmod 4$$

Solution.

Case 1. p **is prime and** $p = 2$**.** Then $-1 \bmod 2 = 1 \mod 2$. We have that $1^2 \equiv 1 \pmod p$. Thus -1 as in $-1 \bmod 2$ is q.r. mod 2.

Case 2. p **is an odd prime.** If $p = 4k+1$ then by Euler's criterion, Eq.(10.110) we have

$$a^{\frac{p-1}{2}} \equiv (-1)^{(p-1)/2} \equiv (-1)^{2k} \equiv 1 \pmod p.$$

Alternatively, $(p-1)/2$ must be even. Thus -1 is a q.r.

For a $p = 4k+3$ we conclude -1 is a q.nr by Euler's criterion as.

$$a^{\frac{p-1}{2}} \equiv (-1)^{(p-1)/2} \equiv (-1)^{2k+1} \equiv -1 \pmod p.$$

■

Exercise 984.
Show the following for an odd prime number $p > 2$.

$$\sum_{a=0}^{p-1} \left(\frac{a}{p}\right) = 0.$$

This is the Legendre symbol indicated.

Solution.
Half of the integers $1, \ldots, p-1$ are q.r. mod p and thus

$$\left(\frac{a}{p}\right) = 1$$

and half are q.nr. mod p and thus

$$\left(\frac{a}{p}\right) = -1$$

For $a = 0$ we have $p|a$ and thus

$$\left(\frac{a}{p}\right) = 0.$$

The number of 1s is equal to the number of -1. ■

Exercise 985.

Let $p \equiv 3 \pmod 4$ and a prime number. Let a be a q.r. mod p. Show that $b = a^{\frac{p+1}{4}}$ is a square root of a mod p.

Solution.
Since $p \equiv 3 \pmod 4$ the $p = 4k+3$ and $p+1 = 4k+4$ and $(p+1)/4 = k+1$ is an integer. Let $b = a^{\frac{p+1}{4}}$. We have the following then.

$$b = a^{\frac{p+1}{4}} \Rightarrow b^2 = a^{\frac{p+1}{2}} = a^{1+\frac{p-1}{2}} \equiv a \pmod p,$$

since a is a quadratic residue mod p and thus

$$\left(\frac{a}{p}\right) = 1 \equiv a^{\frac{p-1}{2}} \pmod p,$$

that is a is a quadratic residue mod p and b is the square root of a or equivalently $b^2 \equiv a \pmod p$. ■

Exercise 986.
(a) Show the following for a p that is an odd prime number.

$$\left(\frac{a}{p}\right)=\left(\frac{b}{p}\right)=1 \quad \Rightarrow \quad \left(\frac{ab}{p}\right)=1.$$

(b) Show the following for a p that is an odd prime number.

$$\left(\frac{a}{p}\right)=\left(\frac{b}{p}\right)=-1 \quad \Rightarrow \quad \left(\frac{ab}{p}\right)=1.$$

Solution.
(a) It has been proven earlier (e.g. previous problem part (c)).
(b) $\mathbb{U}_p$ has a primitive root g and therefore

$$a=g^k \pmod p, \quad b=g^l \pmod p, \quad \left(\frac{a}{p}\right)=-1=\left(\frac{b}{p}\right).$$

We conclude that k and l are odd numbers since otherwise $k/2, l/2$ is an integer and thus $g^{k/2}$ is such that

$$(g^{k/2})^2 \equiv g^k \equiv a \pmod p,$$

and likewise $g^{l/2}$ is such that

$$(g^{l/2})^2 \equiv g^l \equiv b \pmod p,$$

which would implye that a,b are q.r. mod p contradicting the corresponding assumptions i.e. $\left(\frac{a}{p}\right)=-1=\left(\frac{b}{p}\right)$. Thus for odd k,l we have that $(k+l)/2$ is an integer Therefore

$$(g^{\frac{k+l}{2}})^2 \equiv g^k g^l \equiv ab \pmod p,$$

implies that ab is a q.r. mod p. ■

10.21.6 Jacobi warmup

Exercise 987.
Let $m,n \in \mathbb{Z}$ such that $\gcd(m,m) = 1$. Show the following result.

$$a \text{ is q.r. mod } mn \quad \Leftrightarrow \quad a \text{ is q.r. mod } m \quad \wedge \quad b \text{ is q.r. mod } n. \tag{10.111}$$

Solution.
$\Rightarrow$. Let a be a q.r. mod mn. Then there exists an x such that

$$x^2 \equiv a \pmod{mn} \Rightarrow mn | x^2 - a.$$

The latter implies $m|x^2 - a$ and $n|x^2 - a$ since $\gcd(m,m) = 1$. This proves the $\Rightarrow$ part.
Let a be q.r. mod m and n.respectively. Then, there exist x,y such that

$$x^2 \equiv a \pmod{m} \quad \wedge \quad y^2 \equiv a \pmod{n},$$

and therefore there exist M,N integer such that $x^2 - a = mM$ and $y^2 - a = nN$. By the extension of Euclid's theorem by way of $\gcd(m,m) = 1$ we have integer A,B such that

$$Am + Bn = 1$$

Then we derive the following.

$$\begin{aligned} Am + Bn &= 1 \Rightarrow \\ Am(x-y) + Bn(x-y) &= x - y \Rightarrow \\ Cm + Dn &= x - y \Rightarrow \\ Dn + y &= x - Cm \Rightarrow \end{aligned}$$

where $C = A(x-y)$ and $D = B(x-y)$. Consider now $X = x - Cm$ and $Y = y + Dn$. They are $X = Y$. Furthermore,

$$X^2 = (x - Cm)^2 \equiv x^2 \equiv a \pmod{m},$$

and

$$Y^2 = (y + Dn)^2 \equiv y^2 \equiv a \pmod{n}.$$

Since $X = Y$ and thus $X^2 = Y^2$ we have shown that

$$m | X^2 - a \quad \wedge \quad n | X^2 - a,$$

and since $\gcd(m,n) = 1$, this implies

$$nm | X^2 - a,$$

i.e. $X^2 \equiv a \pmod{mn}$, and thus a is q.r. mod mn as needed. ■

Exercise 988.

Let $n = pq$ where p, q are prime numbers and $p \neq q$. Show the following.
If a is a quadratic residue mod n, then a has four square roots in $\mathbb{Z}_p^x$ and thus one quarter of the elements of $\mathbb{Z}_p^x$ are quadratic residues mod n.

Solution.
Let $\mathbb{QR}_n$ be the set of quadratic residues mod n. We define function $\mathbb{Z}^x \mapsto \mathbb{QR}_n$ with $x \mapsto x^2 \pmod{n}$.
If $a \in \mathbb{QR}_n$ then there exists an x such that

$$x^2 \equiv a \pmod{n}.$$

Then since $n = pq$ as defined, we also have the following.

$$x^2 \equiv a \pmod{p},$$

$$x^2 \equiv a \pmod{q}.$$

Thus x is also a square root of a mod p and mod q. Consider now a b, c such that

$$b^2 \equiv a \pmod{p},$$

$$c^2 \equiv a \pmod{q},$$

i.e. one of the square roots of a mod p, and one of the square roots of a mod q. By the Chinese Remainder Theorem, there is a A such that

$$A \equiv B \pmod{p},$$

$$A \equiv C \pmod{q},$$

which imply

$$A^2 \equiv a \pmod{pq}$$

since $\gcd(p, q) = 1$. (See also previous problem.)
Integer a has two square roots mod p. Integer a has two square roots mod q. Thus combining the choices to denote B, and C we have four possibilities for a pair (B, C) above. Thus a has four square roots mod $pq = n$. ■

10.21.7 Jacobi symbol properties

Exercise 989.

Jacobi symbol properties. Let $n,m \in \mathbb{Z}$, $n>0, m>0$ be odd (positive) integers. Furthermore let $N,M \in \mathbb{Z}$ such that $\gcd(nm,NM)=1$. Then the following apply.

$$(a)\left(\frac{N}{n}\right)\left(\frac{N}{m}\right)=\left(\frac{N}{nm}\right),$$

$$(b)\left(\frac{N}{n}\right)\left(\frac{M}{n}\right)=\left(\frac{NM}{n}\right),$$

$$(c)\left(\frac{N^2}{n}\right)=\left(\frac{N}{n^2}\right)=1,$$

$$(d)\left(\frac{N^2M}{n^2m}\right)=\left(\frac{M}{m}\right),$$

and if $N \equiv M \pmod{n}$, then

$$(e)\left(\frac{N}{n}\right)=\left(\frac{M}{n}\right).$$

Solution.
(a) Write n,m as a product of odd primes using the unique factorization theorem.

$$n=p_1p_2\dots p_k, \quad m=q_1q_2\dots q_l.$$

Then we have the following

$$\left(\frac{N}{n}\right)\left(\frac{N}{m}\right)=\left(\frac{N}{p_1}\right)\left(\frac{N}{p_2}\right)\dots\left(\frac{N}{p_k}\right)\left(\frac{N}{q_1}\right)\left(\frac{N}{q_2}\right)\dots\left(\frac{N}{q_k}\right)=\left(\frac{N}{p_1p_2\dots p_kq_1q_2\dots q_l}\right)=\left(\frac{N}{nm}\right).$$

(b) Write n as a product of odd primes using the unique factorization theorem.

$$n=p_1p_2\dots p_k.$$

Then

$$\left(\frac{N}{n}\right)\left(\frac{M}{n}\right)=\left(\frac{N}{p_1}\right)\left(\frac{N}{p_2}\right)\dots\left(\frac{N}{p_k}\right)\left(\frac{M}{p_1}\right)\left(\frac{M}{p_2}\right)\dots\left(\frac{M}{p_k}\right)=\left(\frac{NM}{p_1}\right)\left(\frac{NM}{p_2}\right)\dots\left(\frac{NM}{p_k}\right)=\left(\frac{NM}{p}\right).$$

by way of $x^2 \equiv N \pmod{n}$ and by way of $y^2 \equiv M \pmod{n}$ we have $(xy)^2 \equiv NM \pmod{n}$ and thus for example

$$\left(\frac{N}{p_1}\right)\left(\frac{M}{p_1}\right)=\left(\frac{NM}{p_1}\right).$$

(c) Write n as a product of odd primes using the unique factorization theorem.

$$n=p_1p_2\dots p_k.$$

Given that each p_i is an odd prime using the Legendre symbol

$$\left(\frac{N^2}{p_i}\right)=1,$$

for all $i = 1, \ldots, k$. Therefore the following applies.

$$\left(\frac{N^2}{n}\right) = \left(\frac{N^2}{p_1}\right)\left(\frac{N^2}{p_2}\right)\ldots\left(\frac{N^2}{p_k}\right) = 1.$$

But

$$\left(\frac{N}{p_1}\right)\left(\frac{N}{p_1}\right) = 1 = \left(\frac{N}{p_1^2}\right),$$

for all $i = 1, \ldots, k$. Then we have the following.

$$\left(\frac{N}{n^2}\right) = \left(\frac{N}{p_1}\right)\left(\frac{N}{p_1}\right)\left(\frac{N}{p_2}\right)\left(\frac{N}{p_2}\right)\ldots\left(\frac{N}{p_k}\right)\left(\frac{N}{p_k}\right) = 1 \cdot 1 \cdot \ldots \cdot 1 = 1.$$

(d)

$$\begin{aligned}
\left(\frac{N^2M}{n^2m}\right) &= \left(\frac{N^2}{n^2m}\right)\left(\frac{M}{n^2m}\right) \\
&= 1 \cdot \left(\frac{M}{n^2m}\right) \\
&= \left(\frac{M}{n^2}\right)\left(\frac{M}{m}\right) \\
&= 1 \cdot \left(\frac{M}{m}\right) \\
&= \left(\frac{M}{m}\right).
\end{aligned}$$

(e) If $N \equiv M \pmod{n}$, since $n = p_1 \ldots p_k$ we have $N \equiv M \pmod{p_i}$ for every $i = 1, \ldots, k$. Then using Legendre symbol properties as p_i are prime and odd, $\left(\frac{N}{p_i}\right) = \left(\frac{M}{p_i}\right)$ for all i. We then conclude

$$\left(\frac{N}{n}\right) = \left(\frac{N}{p_1}\right)\left(\frac{N}{p_2}\right)\ldots\left(\frac{N}{p_k}\right) = \left(\frac{M}{p_1}\right)\left(\frac{M}{p_2}\right)\ldots\left(\frac{M}{p_k}\right) = \left(\frac{M}{n}\right),$$

as needed. ■

Exercise 990.

Jacobi symbol properties. Let $n,m \in \mathbb{Z}$, $n>0, m>0$ be odd (positive) integers. Then the following apply.

(a) if $a \equiv b \pmod{n}$ then $\left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$.

(b) $\left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}$.

(c) $\left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}$.

(d) For $\gcd(n,m)=1$

$$\left(\frac{n}{m}\right)\left(\frac{m}{n}\right) = (-1)^{\frac{n-1}{2}\frac{m-1}{2}}.$$

Solution.

(a) The result holds for a Legendre symbol i.e. when n is an odd prime number. In our case n is an odd integer. If it is a prime number it holds by way of the Legendre symbol property. Otherwise it is the product of odd prime numbers, and let $n = p_1 \dots p_k$. Moreover $a \equiv b \mod n$ implies $a \equiv b \mod p_i$ for all $i = 1, \dots, k$. Then from a Legendre property $\left(\frac{a}{p_i}\right) = \left(\frac{b}{p_i}\right)$. Then

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\cdots\left(\frac{a}{p_k}\right) = \left(\frac{b}{p_1}\right)\left(\frac{b}{p_2}\right)\cdots\left(\frac{b}{p_k}\right) = \left(\frac{b}{n}\right).$$

(b) Again recalling Legendre symbol properties, the result holds for a Legendre symbol i.e. when n is an odd prime number. In our case n is an odd integer. If it is a prime number it holds by way of the Legendre symbol property. Otherwise it is the product of odd prime numbers, and let $n = p_1 \dots p_k$.

$$\begin{aligned}
\left(\frac{-1}{n}\right) &= \left(\frac{-1}{p_1}\right)\left(\frac{-1}{p_2}\right)\cdots\left(\frac{-1}{p_k}\right) \\
&= (-1)^{\frac{p_1-1}{2}}(-1)^{\frac{p_2-1}{2}}\dots(-1)^{\frac{p_k-1}{2}} \\
&= (-1)^{\frac{p_1-1}{2}+\frac{p_2-1}{2}+\ldots+\frac{p_k-1}{2}} \\
&= (-1)^{\frac{p_1-1}{2}+\frac{p_2-1}{2}+\ldots+\frac{p_k-1}{2} \bmod 2} \\
&= (-1)^{\frac{p_1 p_2 \cdots p_k - 1}{2}} \\
&= (-1)^{\frac{n-1}{2}}.
\end{aligned}$$

(c) Similarly to part (b). Use induction with base case $n = p_1 p_2$ so that a claim similar to the one used above can be shown i.e.

$$\frac{p_1^2-1}{8} + \frac{p_2^2-1}{8} \bmod 2 = \frac{p_1^2 p_2^2 - 1}{8} = \frac{n^2-1}{8}.$$

(d) See a follow-up problem. ■

Exercise 991.

(a) Let p,q be odd integer numbers such that $\gcd(p,q)=1$. Show the following.

$$\left(\frac{p}{q}\right)=\prod_i\prod_j\left(\frac{p_i}{q_j}\right), \tag{10.112}$$

where $p=p_1p_2\dots p_k$, $q=q_1q_2\dots q_l$, for $k\geq i\geq 1$, $l\geq j\geq 1$ and p_i,q_j prime numbers.
Note: The symbol $\left(\frac{p}{q}\right)$ is the Jacobi symbol as q might not be a prime number and in fact it is a compositive number in the general case. The symbol $\left(\frac{p_i}{q_j}\right)$ is a Legendre symbol though as all q_j are prime numbers.
(b) Then show the following.

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\frac{q-1}{2}}. \tag{10.113}$$

Solution.
Note that $p_i\neq q_j$ since $\gcd(p,q)=1$.

$$\begin{aligned}\left(\frac{p}{q}\right) &= \left(\frac{p_1p_2\dots p_k}{q}\right)=\prod_{i=1}^k\left(\frac{p_i}{q}\right)=\prod_{i=1}^k\left(\frac{p_i}{q_1q_2\dots q_l}\right)\\ &= \prod_{i=1}^k\prod_{j=1}^l\left(\frac{p_i}{q_j}\right).\end{aligned}$$

Likewise one can show the following.

$$\left(\frac{q}{p}\right) = \prod_{i=1}^k\prod_{j=1}^l\left(\frac{q_j}{p_i}\right).$$

Later on, these two results are multiplied together, and the following is derived.

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \prod_{i=1}^k\prod_{j=1}^l\left(\frac{p_i}{q_j}\right)\left(\frac{q_j}{p_i}\right),$$

and the quadratic reciporicity result is derived using Legendre symbol properties and in particular

$$\left(\frac{p_i}{q_j}\right)\left(\frac{q_j}{p_i}\right)=(-1)^{\frac{p_i-1}{2}\frac{q_j-1}{2}}\Leftrightarrow\left(\frac{p_i}{q_j}\right)=\left(\frac{q_j}{p_i}\right)(-1)^{\frac{p_i-1}{2}\frac{q_j-1}{2}}.$$

and then

$$\left(\frac{p}{q}\right)=\prod_{i=1}^k\prod_{j=1}^l\left(\frac{p_i}{q_j}\right)=\prod_{i=1}^k\prod_{j=1}^l\left(\frac{q_j}{p_i}\right)(-1)^{\frac{p_i-1}{2}\frac{q_j-1}{2}}=\left(\frac{q}{p}\right)\prod_{i=1}^k\prod_{j=1}^l(-1)^{\frac{p_i-1}{2}\frac{q_j-1}{2}}=\left(\frac{q}{p}\right)(-1)^{\sum_i\sum_j\frac{p_i-1}{2}\frac{q_j-1}{2}},$$

and then utilizing the previous problem we have.

$$\begin{aligned}\sum_i\sum_j\frac{p_i-1}{2}\frac{q_j-1}{2} &= \sum_i\frac{p_i-1}{2}\sum_j\frac{q_j-1}{2}\\ &= \frac{p_1p_2\dots p_k}{2}\frac{q_1q_2\dots q_l}{2}\\ &= \frac{p-1}{2}\frac{q-1}{2}\pmod 2.\end{aligned}$$

■

Exercise 992.

Function Jacobi(a,n) calculates

$$\left(\frac{a}{n}\right)$$

in polynomial time.

Solution.
We distinguish multiple cases.
Case 1. Say a is even. Then $a = 2^k b$, where b is odd. Therefore we use the following

$$\left(\frac{a}{n}\right) = \left(\frac{\frac{a}{2}}{n}\right)\left(\frac{2}{n}\right) = \ldots \left(\frac{b}{n}\right) \cdot \left(\left(\frac{2}{n}\right)\right)^k.$$

Case 2. If a is odd and $a < n$ use $a \equiv b \pmod{n} \Rightarrow \left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$ and for a,n such that $\gcd(a,n) = 1$ $\left(\frac{a}{n}\right)\left(\frac{n}{a}\right) = (-1)^{\frac{n-1}{2}\frac{a-1}{2}}$, to obtain

$$\left(\frac{a}{n}\right) = \ldots \left(\frac{n'}{a'}\right), n' > a',$$

for some $n' > a'$ dependent on a,n, and this becomes Case 3.
Case 3. If a is odd and $a > n$

$$\left(\frac{a}{n}\right) = \left(\frac{a \bmod n}{n}\right),$$

and continue until a is 1 or 2. ■

10.21.8 Gauss lemma

Exercise 993.

Gauss's Lemma version 1. Let $p>2$ be an odd prime number number. Let $a\in\mathbb{Z}_p^*$ be such that $a>0$ and $\gcd(a,p)=1$.

Consider

$$I(a)=\{1\le i\le (p-1)/2 : ia \pmod p\}.=\{1\le i\le (p-1)/2 : R(i)\},$$

and

$$T=\{1\le i\le (p-1)/2 : R(i)<0\},$$

where $R(i)$ is the absolute least residue of $ai \pmod p$, and thus $-(p-1)/2\le r(i)\le (p-1)/2$, with

$$R(i)\equiv i\cdot a\equiv a(i)r(i) \pmod p,$$

and $a(i)\equiv\pm 1 \pmod p$.

More Legendre symbol properties. Then show that

$$\left(\frac{a}{p}\right)=(-1)^{|T|}.$$

Solution.

Diversion first. In the remainder, $1\le i,j\le (p-1)/2$ and thus $2\le i+j\le p-1$, and $i-j$ or $j-i$ is at least $-(p-1)/2$ and at most $(p-1)/2$. Consider set A and set B defined as follows.

$$A=\{1,2,\ldots,\frac{p-1}{2}\},$$

$$B=\mathbb{Z}_p^*-A=\{\frac{p-1}{2}+1,\frac{p-1}{2}+2,\ldots,p-1\}=\{-1,-2,\ldots,-\frac{p-1}{2}\}.$$

Let $a\in\mathbb{Z}_p^*$ per problem's statement. Consider the set

$$I(a)=a\cdot A=\{a,2a,\ldots,\frac{p-1}{2}a\} \pmod p.$$

For every $i\in A$, i.e. $1\le i\le (p-1)/2$ we have the following

$$a\cdot i\equiv R(i)=a(i)\cdot r(i) \pmod p$$

where $a(i)=\pm 1$ and $r(i)$ defines a mapping $A\mapsto A$ that is an injective function (injection) and $r(i)\equiv r(j) \pmod p \Rightarrow i\equiv j \pmod p$ (or equivalently, $i\not\equiv j \pmod p \Rightarrow r(i)\not\equiv r(j) \pmod p$).

The $r(i)\equiv r(j) \pmod p \Rightarrow i\equiv j \pmod p$, is derived by the following observations, noting that if $r(i)\equiv r(j) \pmod p$, then $ai\equiv aj \pmod p$, or $ai\equiv -aj \pmod p$

Observation 1: $ai\equiv aj \pmod p$ **implies** $i=j$**.**

If $R(i)\equiv R(j) \pmod p$ with $ai\equiv aj \pmod p$, then $i\equiv j \pmod p$ since $ai\equiv aj \pmod p$ implies $(i-j)a=kp$ for some k and since $\gcd(p,a)=1$ we have $p|i-j$; for the range of i,j this can be true only for $i=j$.

Observation 2: $ai\not\equiv -aj \pmod p$**.** If $ai\equiv -aj \pmod p$ then $a(i+j)=kp$ for some k and since $\gcd(p,a)=1$ we have $p|i+j$, We know however that $i+j\le p-1$ and thus $p\nmid i+j$.

Therefore $r(.)$ is a permutation of A. We then have the following

$$\begin{aligned}
a\cdot i &\equiv a(i)r(i) \pmod p \Leftrightarrow\\
\prod_i a\cdot i &\equiv \prod_i a(i)r(i) \pmod p \Leftrightarrow\\
a^{\frac{p-1}{2}}\left(\frac{p-1}{2}\right)! &\equiv \prod_i a(i)\prod_i r(i) \pmod p \Leftrightarrow\\
a^{\frac{p-1}{2}}\left(\frac{p-1}{2}\right)! &\equiv \prod_i a(i)\left(\frac{p-1}{2}\right)! \pmod p.
\end{aligned}$$

Since $\gcd(p, \frac{p-1}{2}!) = 1$ and by Euler's criterion $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$, we obtain the following.

$$\begin{aligned} a^{\frac{p-1}{2}} \left(\frac{p-1}{2}\right)! &\equiv \prod_i a(i) \left(\frac{p-1}{2}\right)! \pmod{p} \Leftrightarrow \\ \left(\frac{a}{p}\right) &= \prod_i a(i) \pmod{p} \Leftrightarrow \left(\frac{a}{p}\right) = (-1)^{|T|} \bmod p \Leftrightarrow \left(\frac{a}{p}\right) = (-1)^{|T|}. \end{aligned}$$

■

Exercise 994.

Gauss's Lemma version 2 . Let $p>2$ be an odd prime number. Let $a\in\mathbb{Z}_p^*$ be such that $a>0$ and $\gcd(a,p)=1$. Consider

$$I(a)=\{1\leq i\leq (p-1)/2 : ia \pmod p\}.=\{1\leq i\leq (p-1)/2 : R(i)\},$$

and

$$T=\{1\leq i\leq (p-1)/2 : R(i)<0\},$$

where $R(i)$ is the absolute least residue of $ai \pmod p$, and thus $-(p-1)/2\leq r(i)\leq (p-1)/2$, with

$$R(i)\equiv i\cdot a\equiv a(i)r(i) \pmod p,$$

and $a(i)\equiv\pm 1 \pmod p$.

More Legendre symbol properties. Then show that

$$\left(\frac{a}{p}\right)=(-1)^{|T|}.$$

Example. A proof by example follows. Let $a=2\in\mathbb{Z}_7^*$ and thus $p=7$.

$$I(a)=\{2\cdot 1,2\cdot 2,2\cdot 3\}=\{2,4,6\}=\{2,4-7,6-7 \pmod 7\}=\{2,-3,-1 \pmod 7\}$$

$$T=\{-3,-1 \pmod 7\}=\{R(i):R(i)<0,1\leq i\leq \frac{p-1}{2}\}$$

and $|T|=2$. Furthermore,

$$\prod_{x\in I(a)} R(i)=\prod_{1\leq i\leq \frac{p-1}{2}} ai=a^{\frac{p-1}{2}}\cdot\frac{p-1}{2}!=2^3\cdot 3!==2^{\frac{p-1}{2}}\cdot 3!=\left(\frac{2}{p}\right)\cdot 3!.$$

Moreover,

$$\prod_{x\in I(a)} R(i)\equiv\prod_{x\in I(a)} a(i)r(i)\equiv(-1)^{|T|}\frac{p-1}{2}!=(-1)^2\cdot 3!=(-1)^{|T|}\cdot 3!$$

Equating the two products and cancelling $\frac{p-1}{2}!$ since $\gcd(p,\frac{p-1}{2}!)=1$, the result follows.

Solution.

Generalize the proof by example argument. It is a less formal proof (or summary) of the previous one. Pick two elements of $I(a)$. If two multiples of a say ia and ja, $1\leq i\neq j\leq (p-1)/2$, are congruent $\pmod p$ i.e. $ia\equiv ja \pmod p$ then $i\equiv j \pmod p$, which leads to $i=j$. Likewise if $ia\equiv -ja \pmod p$ then $i\equiv -j \pmod p$, or $(i+j)\equiv 0 \pmod p$, which leads to $p|i+j$ which is impossible since p is prime and $i+j\leq p-1<p$. Thus the absolute values of the multiples should be distinct. That is $|a|,|2a|,\ldots,|(p-1)/2a|$ are distinct, and form a permutation of $1,\ldots,\frac{p-1}{2}$. Multiplying the multiples the first way we get $a^{(p-1)/2}((p-1)/2)!$. Multiplying them together the other way we get $(-1)^{|T|}((p-1)/2)!$. Equating the two we get $a^{(p-1)/2}\equiv(-1)^{|T|} \pmod p$, i.e. $\left(\frac{a}{p}\right)=(-1)^{|T|}$.

■

10.21.9 Eisenstein theorem and more Legendre symbol properties

Exercise 995.

More Legendre symbol properties.

Eisenstein theorem. If $p>2$ is an odd prime number and $a\neq p$ is odd and thus $\gcd(a,p)=1$, then show

$$\left(\frac{a}{p}\right)=(-1)^{\sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor}, \tag{10.114}$$

using Gauss's Lemma.

The sum of the exponent has a nice geometric interpretation. It is the number of lattice points under the line $y=\frac{a}{p}x$ that are over the x axis between $x=0$ and $x=p/2$.

Solution.

For each $i=1,2,\ldots,(p-1)/2$ we have by the division theorem the following.

$$ia=q(i)p+r(i), \tag{10.115}$$

where $q(i)=\lfloor ia/p\rfloor$ and $0\leq r(i)<p$. If $0\leq r(i)\leq(p-1)/2$ then let $s(i)=r(i)$. If $(p-1)/2<r(i)<p$ then let $s(i)=r(i)-p$ or equivalently $r(i)=s(i)+p$. Then, $R(i)$ of Gauss's Lemma is $R(i)=s(i)$ for all $i=1,2,\ldots,(p-1)/2$. By Gauss's Lemma we obtain the following.

$$\left(\frac{a}{p}\right)=(-1)^{|T|},$$

where $|T|$ is the number of $R(i)<0$ or equivalently the number of $r(i)>(p-1)/2$. We now add-up Eq.(10.115) for all relevant i.

$$\begin{aligned}
\sum_{i=1}^{\frac{p-1}{2}}(ia) &= p\sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}r(i)\Leftrightarrow &(10.116)\\
a\cdot\sum_{i=1}^{\frac{p-1}{2}}i &= p\cdot\sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}s(i)+p|T|\Leftrightarrow\\
a\cdot\sum_{i=1}^{\frac{p-1}{2}}i \pmod 2 &\equiv p\cdot\sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}s(i)+p|T| \pmod 2\Leftrightarrow\\
\sum_{i=1}^{\frac{p-1}{2}}i \pmod 2 &\equiv \sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}s(i)+|T| \pmod 2 &(10.117)
\end{aligned}$$

We note above that a,p are odd and thus $a\equiv 1 \pmod 2$ and $p\equiv 1 \pmod 2$. Furthermore, in Eq.(10.117) the $\sum_i s(i)$ values are $\sum_i(\pm i)$. But mod 2 we have $i\equiv i \pmod 2$ and also $i\equiv -i \pmod 2$. Thus the two sums cancel out mod 2. So we can rewrite Eq.(refeisen2) as follows.

$$\begin{aligned}
\sum_{i=1}^{\frac{p-1}{2}}(ia) &= p\sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}r(i)\Leftrightarrow &(10.118)\\
\sum_{i=1}^{\frac{p-1}{2}}i \pmod 2 &\equiv \sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+\sum_{i=1}^{\frac{p-1}{2}}s(i)+|T| \pmod 2\Leftrightarrow &(10.119)\\
0 &\equiv \sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor+0+|T| \pmod 2\Leftrightarrow\\
|T| &\equiv \sum_{i=1}^{\frac{p-1}{2}}\lfloor\frac{ia}{p}\rfloor \pmod 2. &(10.120)
\end{aligned}$$

We plus the right hand expression of Eq.(10.120) into Gauss's Lemma and the result follows. ■

Exercise 996.

More Legendre symbol properties.

(a) Let p,q be odd primes. Let $p \neq q > 0$. Show the following.

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$$

(b) The following conjectured by Euler was proven by Gauss.

$$\left(\frac{p}{q}\right) = \begin{cases} -\left(\frac{q}{p}\right) & p \equiv q \equiv 3 \pmod 4 \\ \left(\frac{q}{p}\right) & \text{otherwise} \end{cases} \tag{10.121}$$

Equivalently, if either p or q is of the form $4k+1$ then $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$, otherwise $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$.

Solution.

(a) The proof uses Eisenstein theorem and its lattice point interpretation of the sum of its exponent.

Let p,q be odd primes. Let $r = \sum_{i=1}^{(p-1)/2} \lfloor \frac{iq}{p} \rfloor$ be the number of lattice points below $y = \frac{q}{p}x$ and over the x axis and between $x = 0$ and $x = p/2$.

Similarly, let $s = \sum_{i=1}^{(q-1)/2} \lfloor \frac{ip}{q} \rfloor$ be the number of lattice points below $x = \frac{p}{q}y$ and over to the right of the y axis and between $y = 0$ and $y = q/2$.

The line $y = \frac{q}{p}x$ and $x = \frac{p}{q}x$ are the same.

None of the two set of points are double counted as they lie on different areas of the dividing line. No point lies on the line as then $xq = py$ and thus x is a multiple of p and y a multiple of q.

The number of points are all inside the rectangle defined by $x = p/2$ and $y = q/2$ and the two axes. The total number of points is $(p-1)/2 \cdot (q-1)/2$. Thus $r+s = (p-1)/2 \cdot (q-1)/2$.

By Eisenstein's Theorem $\left(\frac{p}{q}\right) = (-1)^r$ and $\left(\frac{q}{q}\right) = (-1)^s$. Thus

$$\left(\frac{p}{q}\right)\left(\frac{q}{q}\right) = (-1)^r(-1)^s = (-1)^{r+s} = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$$

(b) Cases analysis for p,q. If $p = 4k+3$ and $q = 4l+3$, then the product $\frac{p-1}{2}\frac{q-1}{2}$ is an odd number. In any other case, if one or both of p,q is $4k+1$ the product is an even number. The result then follows. ■

Exercise 997.
(a) Calculate,

$$\left(\frac{43}{13}\right).$$

(b) Calculate

$$\left(\frac{13}{43}\right).$$

(c) Calculate for prime $p > 2$, where $p \equiv 1 \pmod 4$ and also $p \equiv 1 \pmod 3$, the following symbol.

$$\left(\frac{3}{p}\right).$$

Solution.
(a) We have a Legendre symbol for 13 is a prime number and odd.

$$\left(\frac{43}{13}\right) = \left(\frac{13\cdot 3+4}{13}\right) = \left(\frac{4}{13}\right) = \left(\frac{2\cdot 2}{13}\right) = \left(\frac{2}{13}\right)\left(\frac{2}{13}\right) = 1.$$

(b) We have a Legendre symbol for 43 is a prime number and odd. Note that $13 = 4\cdot 3+1$, yet $43 = 4\cdot 10+3$. Thus $13 \equiv 1 \pmod 4$, $43 \equiv 3 \pmod 4$. By the previous result

$$\left(\frac{13}{43}\right) = \left(\frac{43}{13}\right) = 1.$$

(c) We show that $\left(\frac{3}{p}\right) = 1$, for p as stated.
Note that $3 = 0\cdot 4+3$. Therefore, $\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right)$ if $p = 4k+1$ and $\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right)$ if $p = 4k+3$. Given that $p \equiv 1 \pmod{}$, this means $\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right)$. Consider $\left(\frac{p}{3}\right)$. The only quadratic residue of 3 is 1 so $\left(\frac{p}{3}\right) = 1$ if $p = 3k+1$. Since $p \equiv 1 \pmod 3$ this is indeed the case. Thus $\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = 1$. ■

Exercise 998.

More Legendre symbol properties.

$\left(\frac{2}{p}\right)$ **version 1.** Show that for p an odd prime >2 the following applies.

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}. \tag{10.122}$$

Note. An Eisenshtein-based proof involving the n-th roots of unity and in particular 8-th roots of unity. $w_8^8 = 1$, $z = \sqrt{i} = w_8$, where $i^2 = -1$, and $z^2 = i$ and $z^4 = i^2 = -1$, and thus $z^8 = w_8^8 = 1$. Moreover $w = z + 1/z$ is such that $w^2 = z^2 + 2 + 1/z^2 = 2 + i - i = 2$, and thus $\sqrt{w} = \sqrt{2}$.

Solution.
By the reciprocity theorem (Euler criterion), we have the following; note that in the third step we utilize the binomial theorem $(a+b)^n$ for $a = z$, $b = 1/z$ and $n = p$.

$$\begin{aligned}\left(\frac{2}{p}\right) &= 2^{\frac{p-1}{2}} = (\sqrt{2})^{p-1} = w^{p-1} \\ &= w^p w^{-1} = (z + \frac{1}{z})^p w^{-1} \\ &\equiv (z^p + z^{-p}) w^{-1} \pmod{p}.\end{aligned}$$

From the exponent $\frac{p^2-1}{8}$ we distinguish the following cases: (1) $p = \pm 1 \pmod 8$ and (2) $p = \pm 3 \pmod 8$.
Case 1: $p = 8k \pm 1$**.** Let $p = 8k+1$. Noting that $z^8 = z^{8k} = 1$ we have the following.

$$z^p + z^{-p} = z^{8k+1} + z^{-8k-1} = z + 1/z = w.$$

Case 1b. Let $p = 8k-1$. Similarly we have the following.

$$z^p + z^{-p} = z^{8k-1} + z^{-8k+1} = 1/z + z = w.$$

We then conclude the following

$$\left(\frac{2}{p}\right) \equiv (z^p + z^{-p}) w^{-1} \equiv w \cdot w^{-1} \equiv 1 \equiv (-1)^{\frac{p^2-1}{8}} \pmod{p},$$

as needed.
Case 2: $p = 8k \pm 3$**.** Let $p = 8k+3$. Noting that $z^8 = z^{8k} = 1$ we have the following.

$$z^p + z^{-p} = z^{8k+3} + z^{-8k-3} = z^3 + 1/z^3 = -w.$$

This is because

$$w = z + 1/z \Rightarrow w^3 = (z + 1/z)^3 = z^3 + 1/z^3 + 3z + 3/z \Rightarrow z^3 + 1/z^3 = w^3 - 3w = w^2 w - 3w = 2w - 3w = -w.$$

The $p = 8k-3$ is proven similarly. We then conclude the following

$$\left(\frac{2}{p}\right) \equiv (z^p + z^{-p}) w^{-1} \equiv -w \cdot w^{-1} \equiv -1 \equiv (-1)^{\frac{p^2-1}{8}} \pmod{p},$$

as needed.

■

Exercise 999.

More Legendre symbol properties.

$\left(\frac{2}{p}\right)$ **version 2.** Show that for $p>2$ an odd prime the following applies.

$$\left(\frac{2}{p}\right)=(-1)^{\frac{p^2-1}{8}} \Leftrightarrow \left(\frac{2}{p}\right)=1 \textbf{ if } p\equiv \pm 1 \pmod 8, \qquad \left(\frac{2}{p}\right)=-1 \textbf{ if } p\equiv \pm 3 \pmod 8. \tag{10.123}$$

Solution.

We use Gauss's Lemma. We note that $a=2$ is such that $\gcd(2,p)=1$ for an odd prime number p. We will then show the result utilizing the fact that

$$\left(\frac{2}{p}\right)=(-1)^{|T|},$$

where

$$T=\{1\leq i\leq (p-1)/2 : R(i)<0\},$$

where $R(i)$ is the absolute least residue of $2i \pmod p$, and thus $-(p-1)/2\leq r(i)\leq (p-1)/2$. Consider $1\leq i\leq (p-1)/2$. If $1\leq 2i\leq (p-1)/2$, then $R(i)>0$ and i is not in T. We thus consider only cases i such that the following holds.

$$\frac{p-1}{2}+1\leq 2i\leq p-1 \quad \Leftrightarrow \quad \frac{p-1}{4}+\frac{1}{2}\leq i\leq \frac{p-1}{2}. \tag{10.124}$$

We will find the values of i that satisfy Eq.(10.124) and thus establish $|T|$. Given that p is a odd number there are four cases two consider

$$p=8k+1, \quad p=8k+3, \quad p=8k+5, \quad p=8k+7.$$

Case 1: $p=8k+1$**.** Eq.(10.124) becomes

$$\frac{p-1}{4}+\frac{1}{2}\leq i\leq \frac{p-1}{2} \Leftrightarrow 2k+1/2\leq i\leq 4k$$

The number of values i that satisfy the latter is $4k-(2k+1)+1=2(2k-1)$ an even number. Thus $(-1)^{|T|}=1$, for an even $|T|=2(2k-1)$.

Case 2: $p=8k+3$**.** Eq.(10.124) becomes

$$\frac{p-1}{4}+\frac{1}{2}\leq i\leq \frac{p-1}{2} \Leftrightarrow 2k+1\leq i\leq 4k+1$$

The number of values i that satisfy the latter is $4k+1-(2k+1)+1=2k+1$ an odd number. Thus $(-1)^{|T|}=-1$, for an odd $|T|=2k+1$.

Case 3: $p=8k+5$**.** Eq.(10.124) becomes

$$\frac{p-1}{4}+\frac{1}{2}\leq i\leq \frac{p-1}{2} \Leftrightarrow 2k+3/2\leq i\leq 4k+2$$

The number of values i that satisfy the latter is $4k+2-(2k+2)+1=2k+1$ an odd number. Thus $(-1)^{|T|}=-1$, for an odd $|T|=2k+1$.

Case 4: $p=8k+7$**.** Eq.(10.124) becomes

$$\frac{p-1}{4}+\frac{1}{2}\leq i\leq \frac{p-1}{2} \Leftrightarrow 2k+2\leq i\leq 4k+3$$

The number of values i that satisfy the latter is $4k+3-(2k+2)+1=2k+2$ an even number. Thus $(-1)^{|T|}=1$, for an even $|T|=2k+2$. The result follows from this four case analysis. ■

Exercise 1000.

More Legendre symbol properties.

$\left(\frac{2}{p}\right)$ **version 3.** Show that for p an odd prime > 2 the following applies.

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} \Leftrightarrow \left(\frac{2}{p}\right) = 1 \textbf{ if } p \equiv \pm 1 \pmod 8, \qquad \left(\frac{2}{p}\right) = -1 \textbf{ if } p \equiv \pm 3 \pmod 8 \tag{10.125}$$

Solution.

We use Gauss's Lemma. We note that $a = 2$ is such that $\gcd(2, p) = 1$ for an odd prime number p. We will then show the result utilizing the fact that

$$\left(\frac{2}{p}\right) = (-1)^{|T|},$$

where

$$T = \{1 \leq i \leq (p-1)/2 : R(i) < 0\},$$

where $R(i)$ is the absolute least residue of $2i \pmod p$, and thus $-(p-1)/2 \leq R(i) \leq (p-1)/2$. Let $Q = \frac{p-1}{2}$. Moreover we distinguish two cases $Q = 2k$ equivalent to $p = 4k+1$ and $Q = 2k+1$ equivalent to $p = 4k+3$. We consider the following product.

$$\begin{aligned} A &= \prod_{1 \leq i \leq (p-1)/2} 2i \\ &= 2 \cdot 4 \cdot 6 \cdot \ldots \cdot Q \cdot (Q+2) \cdot (Q+4) \cdot \ldots \cdot (p-3) \cdot (p-1) \\ &= 2^{\frac{p-1}{2}} \left(\frac{p-1}{2}\right)!. \end{aligned}$$

We then compute A differently. We perform a case analysis depending on whether Q is even or odd.

Case 1: $Q = \frac{p-1}{2} = 2k$ **that is** $p = 4k+1$**.**

Product A then becomes.

$$\begin{aligned} A &= \prod_{1 \leq i \leq (p-1)/2} 2i \\ &= \underbrace{2 \cdot 4 \cdot 6 \cdot \ldots \cdot Q}_{k \text{ terms}} \cdot \underbrace{(Q+2) \cdot (Q+4) \cdot \ldots \cdot (p-3) \cdot (p-1)}_{k \text{ terms}}. \\ &\equiv \underbrace{2 \cdot 4 \cdot 6 \cdot \ldots \cdot Q}_{k \text{ terms}} \cdot \underbrace{(-(Q-1)) \cdot (-(Q-3)) \cdot \ldots \cdot (-3) \cdot (-1)}_{k \text{ terms}} \\ &\equiv \underbrace{2 \cdot 4 \cdot 6 \cdot \ldots \cdot 2k}_{k \text{ terms}} \cdot \underbrace{(-(2k-1) \cdot (-(2k-3)) \cdot \ldots \cdot (-3) \cdot (-1)}_{k \text{ terms}}. \\ &\equiv \underbrace{2 \cdot 4 \cdot 6 \cdot \ldots \cdot 2k}_{k \text{ terms}} \cdot \underbrace{(-1)^k \cdot (2k-1) \cdot (2k-3) \cdot \ldots \cdot (3) \cdot (1)}_{k \text{ terms}}. \\ &\equiv (-1)^k \cdot (2k)! \equiv (-1)^k Q! \\ &\equiv (-1)^k \left(\frac{p-1}{2}\right)! \pmod p. \end{aligned}$$

Equating the two expressions of A leads to the following also using Euler's criterion formula.

$$A = 2^{\frac{p-1}{2}} \left(\frac{p-1}{2}\right)! \equiv \left(\frac{2}{p}\right) \left(\frac{p-1}{2}\right)! \equiv (-1)^k \left(\frac{p-1}{2}\right)! \pmod p.$$

Given that $\gcd(p,((p-1)/2)!)=1$, we have

$$\left(\frac{2}{p}\right) \equiv (-1)^k \pmod{p}.$$

We distinguish subcases for k

Case 1a: $k=2m$**,** $m \in \mathbb{Z}$**.** Then $(-1)^k = 1$ as k is an even number. Reconciling $Q=(p-1)/2=2k=2(2m)$ leads to the $p=8k+1$ case. Then $\left(\frac{2}{p}\right) \equiv (-1)^k \pmod{p} = 1$.

Case 1b: $k=2m+1$**,** $m \in \mathbb{Z}$**.** Then $(-1)^k = -11$ as k is an odd number. Reconciling $Q=(p-1)/2=2k=2(2m+1)$ leads to the $p=8k+5$ case. Then $\left(\frac{2}{p}\right) \equiv (-1)^k \pmod{p} = -1$. For the remaining two cases we consider case 2.

Case 2: $Q=\frac{p-1}{2}=2k+1$ **that is** $p=4k+3$. We work similarly as before and we consider two subcases, Case 2a with $Q=(p-1)/2=2k+1$ with $k=2m$ i.e $p=8m+3$ and Case 2b, with $Q=(p-1)/2=2k+1$ with $k=2m+1$ i.e. $p=8m+7$. The former case leads to $\left(\frac{2}{p}\right) \equiv -1$, and the latter to $\left(\frac{2}{p}\right) \equiv 1$. We can unify the cases analysis to the following one.

$$\left(\frac{2}{p}\right) \equiv (-1)^{\frac{p^2-1}{8}}.$$

■

Exercise 1001.

Consider $p = 11$. Determine the value of $|T|$ of Gauss's Lemma for $a = 2$.

Solution.
Let $a = 2$ and $p = 11$. Then, $Q = \frac{p-1}{2} = 5$.

$$T = \{1 \leq i \leq 5 : R(i) < 0\},$$

where $R(i) = 2i \pmod{11}$, where $R(i)$ is as defined previously. Consider the elements of T or before that

$$2 \cdot 1 = 2, 2 \cdot 2 = 4, 2 \cdot 3 = 6, 2 \cdot 4 = 8, 2 \cdot 5 = 10,$$

of which products 3 are greater than $(p-1)/2 = 5$ and 2 are less than or equal to $(p-1)/2$. Thus $k = 3$. Furthermore

$$\begin{aligned} 2 \cdot 1 &= 0 \cdot 11 + 2 \\ 2 \cdot 2 &= 0 \cdot 11 + 4 \\ 2 \cdot 3 &= 0 \cdot 11 + (11 - 5) = 1 \cdot 11 - 5 \\ 2 \cdot 4 &= 0 \cdot 11 + (11 - 3) = 1 \cdot 11 - 3 \\ 2 \cdot 5 &= 0 \cdot 11 + (11 - 1) = 1 \cdot 11 - 1 \end{aligned}$$

We see that $R(1) = 2$, $R(2) = 4$ but $R(3) = -5$ and $R(4) = -3$ and $R(5) = -1$. Moreover

$$2^{\frac{11-1}{2}} \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \equiv 2^{\frac{11-1}{2}} \frac{11-1}{2}! \equiv a(1)r(1)a(2)r(2)a(3)r(3)a(4)r(4)a(5)r(5) \equiv (-1)^3 \frac{11-1}{2}! \pmod{11}$$

and $a(3) = a(4) = a(5) = -1$ and $a(1) = a(2) = 1$. If we add up the equations we have

$$2 \cdot (1 + 2 + 3 + 4 + 5) = (0 + 0 + 0 + 0 + 0) \cdot 11 + (2 + 4 - 5 - 3 - 1) + 3 \cdot 11$$

If we take $\pmod 2$ both sides having noted that $11 \equiv 1 \pmod 2$ and $2 \equiv 0 \pmod 2$, we obtain the following.

$$(1 + 3 + 5 - 2 - 4) \equiv 3 \pmod 2 \equiv 1 \pmod 2$$

and the right hand size (middle section) is the number of multiples that are negative or equivalently greater than $p/2$ (so that Gauss's Lemma becomes applicable). ■

Exercise 1002.

Let $p > 2$ be an odd prime number. Let $a < p$ or in general $p \nmid a$ and thus $\gcd(p,a) = 1$.
Show that for $k \geq 2$, a is a q.r mod p^k if and only if a is a q.r. mod p.

Solution.
$\Rightarrow$. If a is a q.r. mod p^k, then there exists a b such that $b^2 \equiv a \pmod{p^k}$. That is, $p^k | b^2 - a$. Then $p | b^2 - a$ and the result follows for the forward direction: a is then a q.r $\pmod{p}$ as well.
$\Leftarrow$. For the converse, let $b^2 \equiv a \pmod{p}$. The proof resembles a prior proof and is by induction on k.
Base case. Obviously true for $k = 1$.
Inductive step: from k to $k+1$. Let a be a q.r $\pmod{p^k}$ for $k \geq 1$. We shall show that a is a q.r. $\pmod{p^{k+1}}$.
By the induction hypothesis (inductive assumption) he have $b^2 \equiv a \pmod{p^k}$ i.e. $b^2 - a = p^k r$, for some integer r. Let us form $c = b + dp^k$, where c, d are yet to be determined in full. We observe first that $2k \geq k+1$ for $k \geq 1$. (It will be used to obtain the third derivation below.) We then have the following.

$$\begin{aligned} c^2 - a &\equiv (b + dp^k)^2 - a \\ &\equiv b^2 + 2bdp^k + d^2 p^{2k} - a \pmod{p^{k+1}} \\ &\equiv rp^k + 2bdp^k \\ &\equiv p^k(r + 2bd) \pmod{p^{k+1}} \end{aligned}$$

For $c^2 - a \equiv 0 \pmod{p^{k+1}}$ we need $p | r + 2bd$ in other words $(2b)d \equiv -r \pmod{p}$. Since p is an odd prime $b < p$ and $\gcd(2b, p) = 1$, there is a solution for d of the modular equation $(2b)d \equiv -r \pmod{p}$. Thus $c^2 - a \equiv 0 \pmod{p^{k+1}}$ as needed and this completes the inductive step.
In conclusion if p is a prime and a is a q.r $\pmod{p}$. Let $b^2 \equiv a \pmod{p^k}$ or $r = (b^2 - a)/p^k$, for some integer r. The $c = b + dp^k$ is such that $c^2 \equiv a \pmod{p^{k+1}}$ if and only if c is defined as follows after defining d. Let $t \equiv (2b)^{-1} \pmod{p}$ and thus $d \equiv -rt \pmod{p}$. We have from above $c = b + dp^k = b - \frac{b^2 - a}{p^k} p^k t$, and thus $c^2 \equiv \left(b - t(b^2 - a)\right)^2 \equiv a \pmod{p^{k+1}}$. Thus a square root b of a mod p^k can be used to construct a square root c of a mod p^{k+1}. ■

Exercise 1003.

Let $p > 2$ be an odd prime number. Let $a \in \mathbb{Z}$ with $\gcd(a,p) = 1$. Show that the quadratic congruence

$$x^2 \equiv a \pmod{p^k}$$

(1) either it has NO solution if $\left(\frac{a}{p}\right) = -1$,

(2) or it has two solutions for $\left(\frac{a}{p}\right) = 1$.

Solution.

Note that since $\gcd(a,p) = 1$ it means $nvdpa$ and thus it cannot be $\left(\frac{a}{p}\right) = 0$.

(1) Say $\left(\frac{a}{p}\right) = -1$. We claim the congruence has no solution for x. If it had, and let us call such a solution x, then by the previous problem, the congruence

$$x^2 \equiv a \pmod{p}$$

■

would have a solution, implying $\left(\frac{a}{p}\right) = 1$, a contradiction to $\left(\frac{a}{p}\right) = -1$.

(2) Say $\left(\frac{a}{p}\right) = 1$. The a is a q.r. mod p. By the previous problem a is a q.r. mod p^k. Therefore there exists an x such that

$$x^2 \equiv a \pmod{p^k},$$

and problem completed, noting that if x is a solution then $-x$ is also a solution. By Lagrange's theorem these are the only two solutions.

Exercise 1004.

For $p=5$ find the square root of 11 mod 5^3.

Solution.

Base case $k=1$**.** Let $p=5$. Then $11\equiv 1 \pmod 5$ is obviously a q.r. mod 5. Then 11 is also a q.r mod 5^2 , 5^3 and so on, by the previous problem.

Next step $k=2$**.** We show that 11 is a qudratic residue mod 5^2 by presenting a square root of 11 mod 5^2. We start with the square root of $11 \pmod 5$ mod 5 i.e. $b=1$ and we build into the square root c of 11 mod 5^2 using the previous problem's inductive proof.

Thus from the base case $k=1$ we have $b=1$ a square root mod 11 mod 5 mod 11. We calculate first $r=\frac{b^2-a}{p^1}=\frac{1^2-11}{5}=-2$.

Then we form $c=b+dp^1=b+5d$. For this we need to find first $t=(2\cdot b)^{-1}=(2)^{-1} \pmod 5$. A simple calculation leads that $2\cdot 3\equiv 1 \pmod 5$. Therefore $t=3$.

Then we form $d=-rt=-(-2)3=6\equiv 1 \pmod 5 = 1$.

Therefore $c=b+dp^1=b+5d=1+5\cdot 1=6$.

It is then easy to confirm

$$c^2=6^2\equiv 11 \pmod{5^2}.$$

Nest step $k=3$**.** We show that 11 is a qudratic residue mod 5^3 by presenting a square root of 11 mod 5^3. We start with the square root of $11 \pmod 5)^2$ i.e. $b=6$ (the c of the previous step) and we build into the square root c of 11 mod 5^3 using the previous problem's inductive proof.

Thus from the previous case's $k=2$ we have $b=6$ a square root mod 11 mod 5^2. We calculate first $r=\frac{b^2-a}{p^2}=\frac{6^2-11}{25}=1$.

Then we form $c=b+dp^2=b+25d$. For this we need to find first $t=(2\cdot b)^{-1}=(2\cdot 6)^{-1}=(12)^{-1}\equiv 2^{-1} \pmod 5$. A simple calculation leads that $2\cdot 3\equiv 1 \pmod 5)^2$. Therefore $t=3$.

Then we form $d=-rt=-(1)3=-3\equiv 2 \pmod 5 = 2$.

Therefore $c=b+dp^2=b+25d=6+25\cdot 2=56$. It is easy to confirm then the following.

$$c^2=56^2\equiv 11 \pmod{5^3}.$$

Thus a square root of 11 mod 125 is 56. ■

Exercise 1005.

Show that for $k \geq 3$, $a > 0$ is a q.r mod 2^k if and only if a is an odd integer number and $a \equiv 1 \pmod 8$.
In fact if $a \equiv 1 \pmod 8$ there are four square roots of a mod 2^k: if b is one of them, so is $-b$ mod 2^k, so is $b + 2^{k-1}$ and also $-(b + 2^{k-1})$.

Solution.
Base case $k = 3$**.** We establish the base case $k = 3$ i.e. show that an odd integer a with $a \equiv 1 \pmod 2^3$ is a quadratic residue. For $1, 2, \ldots, 2^3 - 1$, there is only one integer $a = 1$ such that $a \equiv 1 \pmod 8$, and this is obviously $a = 1$. It is straightforward to confirm that 1 is a q.r. mod 8. One has four square roots: 1,3,5, 7 mod 8. We can describe the four square roots as $b = 1 \pmod 8$, $-b \equiv 7 \pmod 8$, $b + 4 = 5$, and $-(b + 4) \equiv 3 \pmod 8$.
Inductive step from 2^k **to** 2^{k+1}**,** $k \geq 3$**.**
By the inductive hypothesis, say a is a q.r mod 2^k, $k \geq 3$ and also $a \equiv 1 \pmod 8$. Then there exists a b such that

$$b^2 \equiv a \pmod{2^k} \tag{10.126}$$

or equivalently $2^k | b^2 - a$, or equivalently $b^2 - a = q2^k$, for some integer k. We will find a square root c of a mod 2^{k+1} then.

$$c^2 \equiv a \pmod{2^{k+1}} \tag{10.127}$$

Case 1: $b^2 \equiv a \pmod{2^{k+1}}$. If it is so we are done: b is a square root of a mod 2^k and also mod 2^{k+1}.
Case 2: $b^2 \not\equiv \pmod{2^{k+1}}$. We have a is odd; this mean b is also odd and thus $b + 1$ is an even number and multiple of two. Just like a prior problem we form c from b as follows

$$c = b + q2^{k-1}.$$

We note that $2k - 2 \geq k + 1$ for $k \geq 3$. Also $b + 1$ is even i.e. $b + 1 = 2B$ for some integer B. We then obtain the following.

$$\begin{aligned}
c &= b + q2^{k-1} \\
c^2 &= (b + q2^{k-1})^2 \\
c^2 - a &= b^2 - a + q^2 2^{2k-2} + 2bq2^{k-1} \\
c^2 - a &\equiv b^2 - a + q^2 2^{2k-2} + 2bq2^{k-1} \pmod{2^{k+1}} \\
c^2 - a &\equiv b^2 - a + 0 + 2bq2^{k-1} \pmod{2^{k+1}} \\
c^2 - a &\equiv 2^k q + 0 + bq2^k \pmod{2^{k+1}} \\
c^2 - a &\equiv 2^k q \cdot (b + 1) \pmod{2^{k+1}} \\
c^2 - a &\equiv 2^{k+1} qB \pmod{2^{k+1}} \\
c^2 - a &\equiv 0 \pmod{2^{k+1}} \\
c^2 &\equiv a \pmod{2^{k+1}}
\end{aligned}$$

A square root mod 2^{k+1} of a has been found. It is straightforward to show that $-c$ is also a square root mod 2^{k+1}. Furthermore,

$$(c + 2^k)^2 = c^2 + 2^{k+1}c + 2^{2k} \equiv c^2 \equiv a \pmod{2^{k+1}},$$

indicates that $c + 2^k$ is a third square root of a, and likewise $-(c + 2^k)$ is a fourth square root of a mod 2^{k+1}. ■

Exercise 1006.
(a) Consider the congruence relation as given below, with a an odd number.

$$x^2 \equiv a \pmod 2$$

Show that it has a solution mod 2.
(b) Consider the congruence relation as given below, with a an odd number.

$$x^2 \equiv a \pmod{2^2}$$

Show that
(b.1) either it has NO solution if $a \equiv 3 \pmod 4$,
(b.2) or it has two solutions if $a \equiv 1 \pmod 4$.
(c) Consider the congruence relation as given below, with a an odd number.

$$x^2 \equiv a \pmod{2^k}$$

Show that for $k \geq 3$
(c.1) either it has NO solution if $a \not\equiv 1 \pmod 8$,
(c.2) or it has four solutions if $a \equiv 1 \pmod 8$; if c is one solution then the other three solutions mod 2^k are $-c$, $c+2^{k-1}$, and $-(c+2^{k-1})$.

Solution.
(a) The only odd number mod 2 is 1. Thus set $a = 1$ It is obvious that there is an $x = 1$ that satisfies the congruence mod 2. The one solution has been found.
(b) If $a \equiv 3 \pmod 4$ it is clear by direct inspection or enumeration that none of $1^2, 2^2, 3^2$ is equivalent to 3 mod 4. If $a \equiv 1 \pmod 4$, we find $x = 1$ and $x = 3$ as two square roots mod 4 and these are the only one by direct inspection.
(c) We use the previous problem. From the base case for $k = 3$ we obtain that only $a \equiv 1 \pmod 8$ has a square root mod 8, and in fact there are four of them: $1, 3, 5, 7$. If $a \not\equiv 1 \pmod 8$ there are no square roots for such a (by inspection). The generalization for $k \geq 3$ is the previous problem. ■

Exercise 1007.

Consider the congruence relation as given below, with a an odd number and $n = 2^{e_1} p_2^{e_2} \dots p_k^{e_k}$, where $e_i \geq 1$, for all $i = 1, 2, \dots, k$, where $p_2, \dots, p_k$ are prime numbers greater than 2.

$$x^2 \equiv a \pmod{n}$$

Show that it has a solution mod n.

Solution.
The following congruences has solutions based on the previous two problems.

$$x^2 \equiv a \pmod{2^{e_1}}$$

$$x^2 \equiv a \pmod{p_i^{e_i}} \quad \forall \quad i = 2, \dots, k.$$

■

By the Chinese Remainder Theorem the problem follows.

Exercise 1008.

There are infinitely many prime numbers p such that $p \equiv 1 \pmod 4$.

Solution.
Say there are only finite prime numbers and let them be $p_1, p_2, \ldots, p_k$, where $p_i \equiv 1 \pmod 4$, $p_i > 2$, $i = 1, \ldots, k$. Then form integer N, and then M, where,

$$N = 2p_1 p_2 \ldots p_k \Rightarrow M = N^2 + 1.$$

M cannot be divisible by any p_i. This is because N and also N^2 are divisible by p_i and that would imply $p_i | M - N^2$ i.e. $p_i | 1$ which contradicts $p_i > 2$. Thus either M is a prime number other than the p_i, and give that $M \equiv 1 \pmod 4$ this is a contradiction, or M is divisible by a prime number q which is $q \equiv 3 \pmod 4$. Then

$$M \equiv 0 \pmod q \Leftrightarrow N^2 + 1 \equiv 0 \pmod q \Leftrightarrow N^2 \equiv -1 \pmod q,$$

where the latter implies that -1 is a quadratic residue mod q, where $q \equiv 3 \pmod 4$ or equivalently $q = 4k + 3$ for some $k \in \mathbb{Z}$. Euler's criterion however contradicts the residuosity claim of -1.

$$\left(\frac{-1}{q}\right) = (-1)^{\frac{q-1}{2}} \equiv (-1)^{\frac{4k+3-1}{2}} \equiv -1,$$

and thus -1 cannot be a quadratic residue mod q. Thus the finiteness assumption is incorrect. ■

Exercise 1009.

Let p be an odd prime number. Let g be a generator of $\mathbb{Z}_p^*$ and let k be an even and positive natural number. Show that g^k is a q.r. if and only if k is an even number.

Solution.
$\Leftarrow$.
Since k is an even number, then $k/2$ is an integer. Therefore

$$g^k \equiv (g^{\frac{k}{2}})^2 \pmod{p},$$

after which we conclude g^k is a quadratic residue mod p.
$\Rightarrow$
If g^k is a qudratic residue then there exists an a such that

$$a^2 \equiv g^k \pmod{p}.$$

Since g is a generator there exists an i such that $a \equiv g^i \pmod{p}$. Therefore

$$g^k \equiv (g^i)^2 \equiv g^{2i} \pmod{p}.$$

Then $k - 2i \equiv 0 \mod \phi(p)$, or equivalently $p-1|k-2i$. Since p is an odd prime, $p-1$ is even. So is $2i$ i.e. k must also be even. ■

Exercise 1010.

Let p be an odd prime number. Let g be a generator of $\mathbb{Z}_p^*$ and let k be an even and positive natural number. As shown earlier, $\mathbb{Z}_p^*$ has $(p-1)/2$ q.r. and thus $\mathbb{Z} = \mathbb{Z}_p^* \cup \{0\}$ has $(p+1)/2$. Since $\mathbb{Z}_p^*$ is cyclic it has a generator g.
(a) Show that g is a q.nr.
(b) Show that $g^2, g^4, \ldots, g^{p-1}$ are q.r and distinct.
(c) Show that $g^1, g^3, \ldots, g^{p-2}$ are q.nr and distinct.

Solution.
(a) Say that g is a q.r. and let a be such that $\gcd(a,p) = 1$ and $g \equiv a^2 \pmod{p}$. It is the case that $a^{p-1} \equiv 1 \pmod{p}$ by Fermat's Little Theorem. Therefore

$$g^{\frac{p-1}{2}} \equiv (a^2)^{\frac{p-1}{2}} \equiv a^{p-1} \equiv 1 \pmod{p}.$$

The latter implies that g can be a generator since $ord_p(g) \leq \frac{p-1}{2}$ instea of $ord_p(g) = \phi(p) = p-1$ as it should be.
(b) It is immediate from the previous problem. The fact that they are distinct is also obvious. Let $g^{2i} \equiv g^{2j} \pmod{p}$ for $2i \neq 2j \leq p-1$, and $i - j < p-1$ leads to $2i - 2j \equiv 0 \pmod{\phi(p)} \equiv 0 \pmod{p-1}$ This implies $p-1 | 2i-2j$ i.e. $p-1 \leq |2i-2j|$. The only possibility for these to happen is $i = j$.
(c) From (b) and the previous problem. By Euler's criterion we know that the number of q.r. is at most $(p-1)/2$ and by (b) it is exactly $(p-1)/2$. Thus the remaining are the q.nr. ■

Exercise 1011.
(a) Is 8 a q.r mod 17? Explain
(b) Are 8, 7, 13 q.r. mod 18? Explain.

Solution.
(a) 17 is an odd prime number. Moreover $5^2 \equiv 8 \pmod{17}$. Thus 8 is a q.r. mod 17, and the/a square of 8 is 5.
(b) 8 is such that $\gcd(8,18) > 1$. Thus 8 is a q.nr. mod 18.
For 7 we have $13^2 \equiv 7 \pmod{18}$. Thus 7 is a q.r. mod 17, For 13 we have $7^2 \equiv 13 \pmod{18}$. Thus 13 is a q.r. mod 17, The square roots of 7 are 13 and 5 mod 18. Note that $23^2 \equiv 5 \pmod{18}$. But $23 \equiv 5 \pmod{18}$. It is thus important to complete our claims with mod 18, as needed. The square roots of 13 are 7 and 11 mod 18. ■

Chapter 11

Primality testing

11.1 Carmichael numbers

Exercise 1012.

Definition 11.1 **Carmichael numbers**

A composite natural number $n > 1$, $n \in \mathbb{N}$, is a Carmichael number if it satisfies the following.

$$a^n \equiv a \pmod{n}, \quad \forall a \in \mathbb{Z}.$$

Definition 11.2 **Carmichael numbers**

[An alternative definition] A composite natural number $n > 1$, $n \in \mathbb{N}$, is a Carmichael number if it satisfies the following.

$$a^{n-1} \equiv 1 \pmod{n}, \quad \forall a \in \mathbb{Z} : \gcd(a,n) = 1.$$

Integer $n = 561$ is a Carmichael number. Show that $n = 561$ is not a prime number by finding an a such that

$$a^{n-1} \not\equiv 1 \pmod{n},$$

a violation of Fermat's Little Theorem is established. How much is

$$a^{n-1} \pmod{n},$$

then?

Solution.
Consider $n = 561$. It is obviously a composite number as n is divisible by 3. Thus n is not a prime number. For a prime number n, $a^{n-1} \equiv 1 \pmod{n}$ for all $a = 1, 2, \ldots, n-1$, by way of Fermat's Little theorem.
Consider $a = 2$. It can be shown that $2^{560} \equiv 1 \pmod{561}$. Obviously then $2^{561} \equiv 2 \pmod{561}$.
Consider now $a = 3$. $3^{560} \equiv 375 \pmod{561} \not\equiv 1 \pmod{561}$, and a violation of Fermat's Little Theorem is obtained for the premise $n = 561$ is a prime number. Note however that $3^{561} \equiv 3 \pmod{561}$. Even though 561 is a Carmichael number and for every a with $\gcd(a, 561) = 1$, we have $a^{560} \equiv 1 \pmod{561}$, note that $a = 3$ has $\gcd(3, 561) = 3 \neq 1$. ■

11.1.1 Korselt's theorem or criterion

Exercise 1013.

Theorem 11.1 **Korselt**

A positive (composite) integer n is a Carmichael number if and only if it is square free and for all prime divisors p of n we have $p-1|n-1$.

Use the Korselt theorem (also known as criterion) to show that all Carmichael numbers are odd.

Solution.
(Using Korselt's theorem.)
Say n is a Carmichael number and it is even. Then $n=2m$, where m is odd since a Carmichael number is square-free and thus m can not be even. Then m has at least one prime factor, let it that be p, and it must be odd (including other ones). We thus have $p|m \Rightarrow p|n$. Then by Korselt's theorem $p-1|n-1$ would imply that an even number $p-1$ divides an odd number $n-1$, a contradiction.
(Another proof without using Korselt's theorem.)
Say n is even. Consider $a=(-1)$ and by Carmichael numbers definition we have the following.

$$(-1)^n \equiv (-1) \pmod{n} \Rightarrow 1 \equiv (-1) \pmod{n} \Rightarrow 2 \equiv 0 \pmod{n}$$

This implies $n|2$. Thus $n \leq 2$. For a positive n this means $n=1$ or $n=2$. A Carmichael number is $n>1$. Thus the only case left is $n=2$. But a Carmichael number is composite and two is prime. Thus n cannot be even, and must be odd. ■

Exercise 1014.
Let $n = p_1 p_2$, $p_1 \neq p_2$, where $p_1, p_2 \in \mathbb{Z}_+$ and prime greater than 2. Let $a \in \mathbb{Z}$ such that

$$a^{n-1} \equiv 1 \pmod{p_1}$$

and

$$a^{n-1} \equiv 1 \pmod{p_2}.$$

Show the following.

$$a^{n-1} \equiv 1 \pmod{p_1 p_2}.$$

The result can be generalized by using induction on n having more than two prime factors.

Solution.
If a is as given then there exist K, L such that

$$a^{n-1} - 1 = K p_1, \quad a^{n-1} - 1 = L p_2.$$

Since p_1, p_2 are prime different from each other $\gcd(p_1, p_2) = 1$, and therefore, there exists x, y integer such that

$$x p_1 + y p_2 = 1.$$

Multiply both sides by $a^{n-1} - 1$. We derive the following.

$$(a^{n-1} - 1)(x p_1 + y p_2) = a^{n-1} - 1 \Rightarrow x L p_1 p_2 + y K p_1 p_2 = a^{n-1} - 1,$$

from which we conclude $a^{n-1} \equiv 1 \pmod{p_1 p_2}$. ■

Exercise 1015.

Prove Korselt's theorem, Theorem 11.1.

Solution.

$\Rightarrow$. We first prove the sufficient condition. Let $n \in \mathbb{Z}_+$, with $n > 1$ that is compositive and a Carmichael number. The following then apply by definition

$$a^{n-1} \equiv 1 \pmod{n}, \quad \forall a \in \mathbb{Z} : \gcd(a,n) = 1,$$

or

$$a^n \equiv a \pmod{n}, \quad \forall a \in \mathbb{Z}.$$

We will then show the following:

(1) n is odd,

(2a) for every prime divisor p of n we have $p^2 \nmid n$, that is, n is square-free, and

(2b) for every prime divisor p of n we have $p-1|n-1$.

Part (1) has been proven in another problem. Set $a = -1$ for example. We then show (2a) and (2b) as follows. Let by the fundamental theorem of arithmetic

$$n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}, \tag{11.1}$$

be the prime decomposition of n, where $a_i \geq 1$ for all $i = 1, \dots, k$. We will show that $a_i = 1$ for all $i = 1, 2, \dots, k$. In the remainder i runs $i = 1, 2, \dots, k$. So does $j \neq i$ if needed. Note that $\gcd(p_i, p_j) = 1$ for all i, j. By prior results on primitive roots, there exist a primitive root b_i mod $p_i^{a_i}$ for all i. Then

$$ord_{p_i^{a_i}}(b_i) = \phi(p_i^{a_i}) = p_i^{a_i - 1}(p_i - 1), \tag{11.2}$$

where $\gcd(b_i, p_i) = 1$. Since $\gcd(p_i, p_j) = 1$ by the C.R.T. we have that there exists an a such that

$$a \equiv b_i \pmod{p_i^{a_i}} \ \forall i \tag{11.3}$$

Note that $\gcd(a, p_i) = 1$ since otherwise $\gcd(b_i, p_i) > 1$, a contradiction to $\gcd(b_i, p_i) = 1$. Since all p_i are prime numbers and relatively prime to each other we have the following by a prior problem.

$$\gcd(a,n) = 1. \tag{11.4}$$

This will be used to derive the second equation below. By the Carmichael number definition we have the following for all $a \in \mathbb{Z}$ and all i.

$$\begin{aligned} a^n &\equiv a \pmod{n} \xrightarrow{\text{by Eq.(11.4)}} \\ a^{n-1} &\equiv 1 \pmod{n} \xrightarrow{p_i^{a_i} | n} \\ a^{n-1} &\equiv 1 \pmod{p_i^{a_i}} \xrightarrow{p_i | n} \\ b_i^{n-1} &\equiv 1 \pmod{p_i^{a_i}} \end{aligned} \tag{11.5}$$

By way of Eq. (11.2) and Eq. (11.5) we obtain the following.

$$p_i^{a_i - 1}(p_i - 1) | n - 1. \tag{11.6}$$

From Eq.(11.1) we have $p_i^{a_i}|n$ and given $a_i \geq 1$, $p_i^{a_i-1}|n$. Furthermore from Eq.(11.6) we have $p_i^{a_i-1}|n-1$. Combining the two we obtain $p_i^{a_i-1}|1$. This is only possible for $a_i = 1$ and this is true for all i. This establishes the square-free consition (2a).
Continuing, $a_i = 1$ and Eq.(11.6 imply $p_i - 1|n-1$ for all i, conditional on $p_i|n$ and this establishes condition (2b). The sufficient part of Korstelt's theorem has been shown.
$\Leftarrow$ We now show the necessary part. Assuming (1), (2a) and (2b) we proceed to show that n is a Carmichael number.
By the fundamental theorem of arithmetic, since $p^2 \nmid n$ for every prime factor p of n we have the following

$$n = p_1 \dots p_k, \tag{11.7}$$

and of course $p_i > 2$ for all $i = 1, \dots, k$. Moreover $p_i - 1|n-1$ by way of condition (2b). We consider two cases.
Case 1. Pick $a \in \mathbb{Z}$ with $\gcd(a, p_i) = 1$. By Fermat's Little Theorem

$$\begin{aligned}
a^{p_i-1} &\equiv 1 \pmod{p_i} \quad \forall i \xLeftrightarrow{p_i-1|n-1:(n-1)=(p_i-1)K} \\
a^{(p_i-1)K} &\equiv 1 \pmod{p_i} \quad \forall i \\
a^{n-1} &\equiv 1 \pmod{p_i} \quad \forall i \xLeftrightarrow{\gcd(a,p_i)=1} \\
a^n &\equiv a \pmod{p_i} \quad \forall i
\end{aligned}$$

Case 2. Pick $a \in \mathbb{Z}$ with $\gcd(a, p_i) = d > 1$. Since $d|a$ and $d|p_i$ for all i, we from the latter $d \leq p_i$ and since $d > 1$ it can only be $d = p_i$ for a prime p_i for all i. Then $p_i|a$ for all i. Trivially then,

$$a^n \equiv p_i^n \equiv p_i \equiv a \equiv 0 \pmod{p_i}$$

Therefore $a^n \equiv a \pmod{p_i}$ forall i using a prior problem leads to $a^n \equiv a \pmod{n}$ since all p_i are relatively prime to each other by way each one b eing a prime number.
An alternative proof of the sufficient condition $\Rightarrow$ follows.
For a Carmichael number n we have for all a $a^n \equiv a \pmod{n}$. Pick a p, such that $p|n$. Then by way of n being a Carmichael number $p^n \equiv p \pmod{n}$, or $n|p^n - p$. Since $p|n$ the prior result shows that $p|p^n - p$. Say $p^2|n$. Then $p^2|p^n - p$ which would imply $p|p^{n-1} - 1$. Obviously $p|p^{n-1}$, $n > 1$. The last two imply $p|1$ and $p = 1$ a contradiction to the primality of p. Therefore $p^2 \nmid n$ for $p|n$.
In order to show that for $p|n$ we have $p-1|n-1$ we start with the definition of a Carmichael number. For $a \in \mathbb{Z}$ we have $a^n \equiv a \pmod{n}$, and for a such that $\gcd(a,n) = 1$ we further have than $a^{n-1} \equiv 1 \pmod{n}$. For any p such that $p|n$ we have as a result $a^{n-1} \equiv 1 \pmod{p}$. Consider a generator a of $\mathbb{Z}_p^*$. We have $a^{p-1} \equiv 1 \pmod{p}$. From this equivalence and the $a^{n-1} \equiv 1 \pmod{n}$ we conclude $(n-1) - (p-1) = k\phi(p) = k(p-1)$ that results into $p-1|n-1$ as needed. ■

Exercise 1016.

Show that the two definitions of Carmichael numbers are equivalent.

Solution.
$\Leftarrow$.
Let

$$a^n \equiv a \pmod{n}, \quad \forall a \in \mathbb{Z}.$$

If $\gcd(a,n) = 1$ we obtain the following.

$$a^n \equiv a \pmod{n} \Rightarrow a(a^{n-1}-1) \equiv 0 \pmod{n} \Rightarrow n|a(a^{n-1}-1) \xRightarrow{\gcd(a,n)=1} n|(a^{n-1}-1) \Rightarrow a^{n-1} \equiv 1 \pmod{n}$$

$\Rightarrow$.
Let n be a Carmichael number. By Korselt's theorem it is squarefree. We need to show $a^n \equiv a \pmod{n}$ for all $a \in \mathbb{Z}$. Since n is squarefree it is the products of mutually distinct primes. It thus suffices to show $a^n \equiv a \pmod{p}$ for all prime divisors p of n.
Case 1: a is divided by p. The $a \equiv 0 \pmod{p}$ and $a^n \equiv a \pmod{p}$ is obviously true.
Case 2: $p \nmid a$. Since p is a prime number we use Fermat's Little theorem. $a^{p-1} \equiv 1 \pmod{p}$. Furthermore by Korselt's theorem $p-1|n-1$ and therefore $a^{n-1} \equiv 1 \pmod{p}$. ■

Exercise 1017.

(a) Show that a Carmichael composite odd number n is the product of at least three prime numbers.
(b) Moreover every prime factor is less than $\sqrt{n}$.

Solution.
(a)
Proof 1. If n is a Carmichael number then by Korselt's theorem it is square free and for all prime divisors p of n we have $p-1|n-1$. All p,q,n below are odd integer numbers.
Let $n=pq$ be the product of just two primes, where $p<q$ and $p\neq q$ since n is squarefree. By Carmichael properties $p-1|n-1$ and $q-1|n-1$.

$$\begin{aligned} q &\equiv 1 \pmod{q-1} \Rightarrow \\ n=pq &\equiv p\cdot 1 \pmod{q-1} \\ n-1 &\equiv p-1 \pmod{q-1} \end{aligned}$$

Since $p<q$ we have $p-1<q-1$. This means that $p-1$ is the remainder of the division of $n-1$ by $q-1$ i.e. $q-1\nmid n-1$ that contradicts the stated $q-1|n-1$.
Proof 2. Reach the same conclusion as before by noting the following

$$\begin{aligned} n &\equiv pq \\ n-1 &\equiv p(q-1)+(p-1) \\ \frac{n-1}{q-1} &\equiv p+\frac{p-1}{q-1} \end{aligned}$$

The $\frac{n-1}{q-1}$ is an integer since $q-1|n-1$, p is obviously an integer, and therefore $\frac{p-1}{q-1}$ must be an integer i.e. $q-1|p-1$. This would imply $q-1\leq p-1$ i.e. $q\leq p$ that contradicts $p<q$.
(b) From the previous case $q-1|p-1$ implies, since $n=pq$, that $q-1|n/q-1$. Therefore $q\leq n/q$. If $q=n/q$ it means $n=q^2$ contradicting n bein square free. Therefore $q<\sqrt{n}$. Furthermore $p<q<\sqrt{n}$ as well. Note that then $pq<n$ and therefore there must be a third prime factor r for $pqr=n$, providing sort of a third proof to (a). ■

11.2 Fermat primality

Exercise 1018.

Lucas theorem. Show that $n > 2$ is a prime number if and only if the following holds for some integer a such that $1 < a < n$.

$$a^{n-1} \equiv 1 \pmod{n} \quad \wedge \tag{11.8}$$

$$a^{\frac{n-1}{p}} \not\equiv 1 \pmod{n}, \tag{11.9}$$

for every prime divisor p of $n-1$.

If no such a exists then n is composite or 1 or 2; the latter two cases are eliminated by way of $n > 2$.

Solution.
$\Rightarrow$ If n is a prime number then the $\mathbb{U}_n$ multiplicative group is a cyclic one with $n-1$ elements. It then has a generator g. Therefore $g^{n-1} \equiv 1 \pmod{n}$, but $g^i \not\equiv 1 \pmod{n}$ for $i < n-1$. Every element a of $\mathbb{U}_n$ can be written in the form $a \equiv g^k \pmod{n}$. Then $a^{n-1} \equiv (g^{n-1})^k \equiv 1 \pmod{n}$. Then $ord_n(a)$ divides $n-1$ and of course $ord_n(a) < n-1$ unless a is a generator. Thus we focus on generators such as g in the remainder. Moreover, consider a prime divisor p of $n-1$. We must have for a generator g, $g^p \not\equiv 1 \pmod{n}$ for every p since otherwise if $g^p \equiv 1 \pmod{n}$, $p \neq n-1$, then we have the following.

$$p|n-1 \Rightarrow \frac{n-1}{p} > 1 \text{and is integer} \Rightarrow g^p \equiv 1 \pmod{p} \Rightarrow ord_n(g) = p < n-1,$$

a contradiction to g being a generator and thus $ord_n(g) = n-1$.
$\Leftarrow$ Consider that there exists an element a such that

$$a^{n-1} \equiv 1 \pmod{n}.$$

Then $ord_n(a)|n-1$. Moreover, $\gcd(a,n) = d$ shows $d = 1$. Otherwise if $d > 1$, $d|a^{n-1}$ and $d|n$ and therefore $d|1$ and $d \leq 1$. This $(\gcd(a,n) = 1)$ implies by Euler's theorem

$$a^{\phi(n)} \equiv 1 \pmod{n},$$

and thus $ord_n(a)|\phi(n)$. Let $ord_n(a) = k$. Then from the first equation above we have the following: $k|n-1$. There are two possibilities: $k = n-1$ and n is prime, or $k < n-1$. For the latter case, $(n-1)/k = pK > 1$ is an integer and either a prime p (and $K = 1$) or the product of a prime p and a composite $K > 1$. Consider then

$$a^{\frac{n-1}{p}}.$$

By the second condition of the problem we have that for every prime divisor of $n-1$ and thus for the prime p

$$a^{\frac{n-1}{p}} \not\equiv 1 \pmod{n}.$$

However we also have the following.

$$a^{\frac{n-1}{p}} \equiv a^{k \cdot K} \equiv (a^k)^K \equiv 1^K \equiv 1 \pmod{n},$$

leading to a contradiction. The only case left is $k = n-1$. If the order of an element is $n-1$ this means n is prime. This is because it implies $n-1|\phi(n)$ and $\phi(n) \leq n-1$, i.e. $\phi(n) = n-1$ and thus n is a prime number. ■

Exercise 1019.

Show the following theorem.

Theorem 11.2

Let $n > 5$ be an odd composite number. The following statements are equivalent.

1. $$\forall a \in \mathbb{Z}, \gcd(a,n) = 1 \Rightarrow a^{\frac{n-1}{2}} \equiv \pm 1 \pmod{n}.$$

2. *n is square-free and for every prime number p such that $p|n$ we have $p-1|\frac{n-1}{2}$.*

$$Composite(n), SquareFee(n), \forall p, Prime(p) \wedge p|n \Rightarrow p-1|\frac{n-1}{2}.$$

Solution.

(1) $\Rightarrow$ (2).

If $\gcd(a,n) = 1$ and $a^{\frac{n-1}{2}} \equiv \pm 1 \pmod{n}$ then $a^{n-1} \equiv 1 \pmod{n}$ and n is a Carmichael number. By Korselt's theorem n is square-free. Moreover for every prime divisor p of n we have by Korselt's theorem also that $p-1|n-1$. We thus need to show that $p-1|\frac{n-1}{2}$ to complete the proof.

Show $p-1|\frac{n-1}{2}$. Let $n-1 = (p-1)q$ by Korselt's theorem and the fact that $p-1|n-1$. There are two possibilies: (a) q is even, or (b) q is odd.

Case (a): q is even. Then $q/2$ is integer and

$$n-1 = (p-1)q \Rightarrow \frac{n-1}{2} = (p-1)\frac{q}{2} \Rightarrow p-1|\frac{n-1}{2},$$

since $q/2$ is an integer.

Case (b): q is odd. We will show that this case is not possible. Therefore case (a) is the only possible case that leads to the desired outcome : $p-1|\frac{n-1}{2}$ as proven.

If q is odd, then $n-1 = (p-1)q$ and $n-1$ is even for odd n, and thus we have the following.

$$a^{\frac{n-1}{2}} \equiv a^{\frac{(p-1)q}{2}} \equiv \left(a^{\frac{p-1}{2}}\right)^q \equiv \left(\left(\frac{a}{p}\right)\right)^q \pmod{p}.$$

Since q is odd, the latter part works as follows.

$$\left(\frac{a^2}{p}\right) = 1 \Rightarrow \left(\left(\frac{a}{p}\right)\right)^q \equiv \left(\frac{a}{p}\right) \pmod{p}$$

By transitivity from the previous derivations we have

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}.$$

For a generator g of $\mathbb{U}_p$ we have $g^{\phi(p)} \equiv g^{p-1} \equiv 1 \pmod{p}$, and also $g^{\frac{p-1}{2}} \equiv -1 \pmod{p}$, by way of Lagrange's theorem. Then, by the Chinese Remainder Theorem for $n = pP$ and noting because n is square-free $\gcd(p,P) = \gcd(p,\frac{n}{p}) = 1$ we can find an a such that the following hold.

$$a \equiv g \pmod{p},$$

and

$$a \equiv 1 \pmod{\frac{n}{p}}.$$

Then

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{p}\right) \equiv \left(\frac{g}{p}\right) \equiv g^{\frac{p-1}{2}} \equiv -1 \pmod{p}. \tag{11.10}$$

Moreover $a \equiv 1 \pmod{\frac{n}{p}}$ and therefore

$$a^{\frac{n-1}{2}} \equiv 1 \pmod{\frac{n}{p}}. \tag{11.11}$$

By way of the consequent of part (1) we have two possibilities: (i) $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$, (ii) $a^{\frac{n-1}{2}} \equiv -1 \pmod{n}$.

Case (b1). $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$**.** We have by part (1) $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$ as one possibility. If it is true then since $p|n$ that $a^{\frac{n-1}{2}} \equiv 1 \pmod{p}$. By way of Eq.(11.10) we then conclude that $p|2$ that leads to $p = 2$ an impossibility since n is odd, and p is a factor of n. The other case left from part (1) is the following one.

Case (b2). $a^{\frac{n-1}{2}} \equiv -1 \pmod{n}$**.** Then considering $\frac{n}{p} = P$, since $P|n$ we then conclude that $a^{\frac{n-1}{2}} \equiv -1 \pmod{P}$. By way of Eq.(11.11) we also conclude that $P|2$ which leads to $P = 2$, an impossibility since n is odd. Therefore case (b) is impossible only case (a) is possible and by case (a) we have already concluded $p-1|\frac{n-1}{2}$. Proof completed.

(2) $\Rightarrow$ (1).

Since n is square free let $n = p_1 p_2 \dots p_k$. By Fermat's little theorem for every a such that $\gcd(a, p_i) = 1$ we have $a^{p_i - 1} \equiv 1 \pmod{p_i}$. Furthermore $p_i|n$ implies $p_i - 1|((n-1)/2)$ or equivalently $(n-1)/2 = (p_i - 1)k_i$ for some integer k_i. Then

$$a^{\frac{n-1}{2}} \equiv a^{(p_i-1)k_i} \equiv (a^{p_i-1})^{k_i} \equiv 1 \pmod{p_i}.$$

By the CRT given that $\gcd(p_i, p_j) = 1$ we have that

$$a^{\frac{n-1}{2}} \equiv 1 \pmod{n}.$$

■

Exercise 1020.
Let $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z} = \{0,1,\ldots,n-1\}$. Let $\mathbb{U}_p$ be the units of $\mathbb{Z}_p$, where p is a prime number greater than two. It is also represented by $\mathbb{Z}_p^x$ or $(\mathbb{Z}/p\mathbb{Z})^x$.
Formally

$$\mathbb{U}_p = \left\{a : a \in \mathbb{Z}_p, a \text{ has an inverse}\right\}.$$

(a) Show that $\mathbb{U}_p$ is a group under multiplication mod p.
The result is true for composite n as well; $\mathbb{U}_n$ is then all i such that $\gcd(i,n) = 1$.
(b) Use part (a) to prove Fermat's little theorem.

Solution.
(a) It suffices to show that $\mathbb{U}_p$ is closed under under multiplication mod p. Consider $a,b \in \mathbb{U}_p$. For prime p and $1 \leq a,b < p$ we have $\gcd(a,p) = \gcd(b,p) = 1$ and thus a^{-1}, b^{-1} exist. Then

$$(ab) \in \mathbb{U}_p,$$

since $(ab)^{-1}$ exists and it is $b^{-1}a^{-1}$ by way of

$$(ab)(b^{-1}a^{-1}) = a(b(b^{-1}a^{-1})) = aa^{-1} = 1,$$

and

$$(b^{-1}a^{-1})(ab) = (b^{-1}(a^{-1}a)b = b^{-1}b = 1.$$

All other properties (e.g. associativity) hold for $\mathbb{U}_p$ by way of holding for $\mathbb{Z}_p$. The result is true for composite n as well; $\mathbb{U}_n$ is then all i such that $\gcd(i,n) = 1$.
(b) If p is a prime number then

$$\mathbb{U}_p = \{1,2,\ldots,p-1\}.$$

If $p \nmid a$ then if $a > p-1$, consider b such that

$$b = a \bmod p \Leftrightarrow a \equiv b \pmod{p}, \text{ where } b \in \{1,2,\ldots,p-1\}.$$

By Lagrange's theorem $ord_p(b) | ord(\mathbb{U}_p)$ i.e. $ord_p(b) | p-1$, or $p-1 = ord_p(b)l$. If $ord_p(b) = k < p-1$, then

$$b^k \equiv 1 \pmod{p} \Rightarrow b^{p-1} = b^{kl} = (b^k)^l \equiv 1 \pmod{p}.$$

Since $a \equiv b \pmod{p}$, we have

$$a^{p-1} \equiv b^{p-1} \equiv 1 \pmod{p},$$

and the result follows. ■

Exercise 1021.

Let $\mathbb{U}_p$ be the units of $\mathbb{Z}_p$, where p is a prime number greater than 2. Let

$$A = \left\{ a \in \mathbb{U}_p : a^{\frac{p-1}{2}} \equiv 1 \pmod{p} \right\}.$$

Show that A is a subgroup of $\mathbb{U}_p$ under modular multiplication mod p.

Solution.

Consider $c, d \in A$. It is easy to show that $cd \bmod p \in A$. Thus the order of A divides the order of $\mathbb{U}_p$ which is $p-1$. Since p is a prime number greater than two, the p is odd and thus $(p-1)/2$ is an integer. We have shown that the order of A is a divisor of $p-1$. We will show that the order of A can either be $p-1$ or $(p-1)/2$ first by showing that A has at least $(p-1)/2$ elements. We can then conclude that the order of A is $(p-1)/2$ by showing that it cannot be $p-1$.

Consider then $b = a^2 \pmod{p}$. We show that $b^2 \in A$. In order to show inclusion to A we must show that $b^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. This is shown as follows.

$$b^{\frac{p-1}{2}} = (a^2)^{\frac{p-1}{2}} = a^{p-1} \equiv 1 \pmod{p},$$

by Fermat's Little theorem as $\gcd(a, p) = 1$ for prime p and $1 \leq a < p$. Furthermore, b is a quadratic residue mod p since there exists an a such that $a^2 \equiv b \pmod{p}$. Thus all quadratic residues are in A. There are $(p-1)/2$ of them. The order of A is thus at least $(p-1)/2$. It can thus be $(p-1)/2$ or $p-1$. We proceed to dismissing $p-1$ as a possibility.

. A generator g of $\mathbb{U}_p$ is not in A since $g^{\phi(p)} = g^{p-1} \equiv 1 \pmod{p}$ and $g^{(p-1)/2} \not\equiv 1 \pmod{p}$. Thus $|A| < p-1$, and therefore it must be $(p-1)/2$ if it can't be $p-1$. Moreover consider $t \notin A$. We have $t^{p-1} \equiv 1 \pmod{p}$ but since $t \notin A$, we have $t^{\frac{p-1}{2}} \not\equiv 1 \pmod{p}$, and it should be $t^{\frac{p-1}{2}} \equiv -1 \pmod{p}$, as the only two square roots of $x^2 \equiv 1 \pmod{}$ by Lagrange's theorem for prime p are $+1$ and -1 mod p. Note that $u = t^{\frac{p-1}{2}}$ is such that $u^2 \equiv 1 \pmod{p}$ and thus if $u \not\equiv 1 \pmod{p}$ then it must be $u = t^{\frac{p-1}{2}} \equiv -1 \pmod{p}$. ■

11.2.1 A Fermat Little Theorem-based primality test

Exercise 1022.

Fermat primality test.

Algorithm 2 is a probabilistic algorithm for testing whether n is composite or prime. Show that the probability that the output of Algorithm 2 is Composite given that n is a composite number but not a Carmichael number is at least $1-2^{-t}$.

Solution.

In other words, if n is a prime number Algorithm 2 returns Pseudoprime with probability one. If n is a composite

Algorithm 2: FermatPrimalityTest

Input: $n > 4$ is odd; number of runs is t
Output: n is composite or pseudoprime

```
1  i = 0 ;
2  do
3  |  i = i + 1 ;
4  |  Pick a ∈ {2,3,...,n−2} uniformly at random ;
5  |  if gcd(a,n) ≠ 1 then
6  |  |  return Composite
7  |  end
8  |  if a^{n−1} ≢ 1 (mod n) then                    /* Compositeness check */
9  |  |  return Composite ;                  /* a is a witness of n's compositeness */
10 |  end
11 while i < t;
12 return PseudoPrime ;                          /* n is either prime or composite */
```

number but not a Carmichael number then Algorithm 2 return composite with probability at least $1-2^{-t}$. Suppose that $\exists a \in \mathbb{U}_n$ such that

$$a^{n-1} \not\equiv 1 \pmod{n}.$$

Consider the following two sets B_n and C_n as follows.

$$B_n = \left\{a : a \in \mathbb{U}_n, a^{n-1} \not\equiv 1 \pmod{n}\right\},$$

$$C_n = \left\{a : a \in \mathbb{U}_n, a^{n-1} \equiv 1 \pmod{n}\right\},$$

We have $|B_n| + |C_n| = \phi(n) = |\mathbb{U}_n|$. For a composite n, we have $|D_n| > 1$ as it contains the prime factors of n, minimally. Note that $a \in \mathbb{U}_n$ implies $\gcd(a,n) = 1$. Set C_n is a subgroup. Consider $k, l \in C_n$ with $k^{n-1} \equiv 1 \pmod{n}$ and $l^{n-1} \equiv 1 \pmod{n}$. We conclude easily that $(kl)^{n-1} \equiv 1 \pmod{n}$ to prove closure. Sidelining the claims about the obvious properties, consider $K = k^{-1}$ the multiplicative inverse of k and consider $L = l^{-1}$ the multiplicative inverse of l. We then have

$$Kk \equiv 1 \pmod{n} \Rightarrow (Kk)^{n-1} \equiv 1 \pmod{n},$$

from which we conclude $K \in C_n$. Similarly for $l \in C_n$ we can show that $L \in C_n$. Moreover, we can show $(kl)^{-1} \in C_n$. Thus C_n is a subgroup of $\mathbb{U}_n$ and the order of C_n divides the order of $\mathbb{U}_n$ which is $\phi(n)$. Thus if $C_n \neq \mathbb{U}_n$ (e.g. n is not a Carmichael number), then $|C_n| \leq |\mathbb{U}_n|/2$ and $|B_n| \geq |\mathbb{U}_n|/2$. ■

11.3 Solovay-Strassen primality test

Exercise 1023.

The primality testing algorithm by Solovay-Strassen uses Euler's property for the Jacobi (composites) or Legendre (primes) symbol rather than Fermat's Little theorem. We know that for a prime number $p > 2$ which is odd we have by Euler's theorem the following holds (Legendre symbol use).

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}, \forall a : 1 \leq a < p.$$

Therefore were we to test if p was a prime number finding an a that violates the condition above proves the following: (a) p is not a prime number, and (b) a is a witness of the compositeness of p. Note that for an odd composite number $n > 2$ the symbol below is a Jacobi symbol.

$$\left(\frac{a}{n}\right) \equiv a^{\frac{n-1}{2}} \pmod{n}, \forall a : 1 \leq a < n.$$

Consider $n = 1387$, $n = 561$ and $n = 63$. Pick $a = 2$. Examine whether $a = 2$ is a Fermat witness and or Euler witness (for the Solovay-Strassen algorithm to follow) of the compositeness of n.

Solution.
(a) $n = 1387 = 19 \cdot 73$**.** Pick $a = 2$, as instructed.
It is

$$a^{1387-1} \equiv 1 \pmod{1387},$$

and

$$a^{\frac{1387-1}{2}} \equiv 51 \pmod{1387}.$$

In the former case a is NOT a Fermat witness; the latter case indicates that a is an Euler witness. This is because $\left(\frac{a}{n}\right)$ should be 1, -1 or 0.
(b) $n = 561 = 3 \cdot 11 \cdot 17$**.** Pick $a = 2$, as instructed.
It is

$$a^{561-1} \equiv 1 \pmod{561},$$

and

$$a^{\frac{561-1}{2}} \equiv 1 \pmod{561}.$$

In the former case a is NOT a Fermat witness; the latter case indicates that a is also NOT an Euler witness.
(c) $n = 63 = 3^2 \cdot 7$**.** Pick $a = 2$, as instructed.
It is

$$a^{63-1} \equiv 4 \pmod{63},$$

and

$$a^{\frac{63-1}{2}} \equiv 2 \pmod{63}.$$

In the former case a is a Fermat witness; the latter case indicates that a is also an Euler witness. ■

Exercise 1024.

Solovay-Strassen. The primality testing algorithm by Solovay-Strassen uses Euler's property/criterion.

Algorithm Solovay-Strassen is a probabilistic algorithm for testing whether n is composite or not. Show that if n is an odd composite number greater than two, then the probability that the output of the algorithm is Composite given that n is a composite number is at least $1-2^{-t}$.

Solution.

The algorithm tests for two conditions a number of times. We denote as t the numer of times these conditions are tested. Every time these conditions are tested we pick a random a from the set $\{2,3,\ldots,n-2\}$ of cardinality $n-1$. The first condition $C_1(a,n)$ checks whether a has a common divisor with n. A positive answer serves makes a a witness of the compositeness of n. We call then a an Euler (test) compositeness witness and denote it by Ew(a,n) or the set of witnesses for n is denoted as Ew(n).

$$C_1(a,n) : \gcd(a,n) > 1.$$

The second condition is Euler's criterion.

$$C_2(a,n) : \left(\frac{a}{n}\right) \not\equiv a^{\frac{n-1}{2}} \pmod{n}.$$

Therefore a becomes a Euler test compositeness witness if and only if the following applies.

$$a \in Ew(n) \Leftrightarrow C_1(a,n) \vee C_2(a,n).$$

By way of $\overline{(C_1(a,n) \vee C_2(a,n))} = \overline{C_1(a,n)} \wedge \overline{C_2(a,n)}$ we obtain the following condition for Euler (test) compositeness non-witnesses Enw(a,n) or the set Enw(n).

$$a \in Enw(n) \Leftrightarrow \overline{C_1(a,n)} \wedge \overline{C_2(a,n)}.$$

The proof of correctness follows from the prior discussion and thus omitted.

Algorithm 3: Solovay-Strassen primality test algorithm

Input: $n > 4$ is odd; number of runs is t
Output: n is composite or pseudoprime

```
1  i = 0 ;
2  do
3  |  i = i + 1 ;
4  |  Pick a ∈ {2,3,...,n−2} uniformly at random ;
5  |  if gcd(a,n) > 1 then                                   /* C1(a) */
6  |  |  return Composite ;          /* a is a witness of n's compositeness */
7  |  end
8  |  Calculate (a/n) , a^((n−1)/2) (mod n);
9  |  if (a/n) ≢ a^((n−1)/2) (mod n) then                    /* C2(a) */
10 |  |  return Composite ;          /* a is a witness of n's compositeness */
11 |  end
12 while i < t;
13 return PseudoPrime ;                  /* n is either prime or composite */
```

Consider the following two sets B_n and C_n as follows, describing in fact Ew(n) and Enw(n).

$$Ew(n) = B_n = \left\{a : a \in \mathbb{U}_n, a^{\frac{n-1}{2}} \not\equiv \left(\frac{a}{n}\right) \pmod{n}\right\},$$

$$Enw(n) = C_n = \left\{a : a \in \mathbb{U}_n, a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}\right\},$$

Consider also set D_n defined as follows.

$$D_n = \{a : 1 \leq a < n \wedge \gcd(a,n) \neq 1\}.$$

We have that if $a \in B_n$ then n is composite and of course $a \in Ew(n)$. We have that if $a \in C_n$ then Solovay-Strassen can't figure out whethe n is prime or composite and of course $a \in Enw(n)$. We prove a sequence of claims to derive our result.

(a) Show that C_n is a subgroup of $\mathbb{U}_n$.

If $a,b \in C_n$ then it is trivial to show that $ab \in C_n$ as well.

$$\begin{aligned} a \in C_n \Rightarrow a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}, \quad \wedge \quad & b \in C_n \Rightarrow b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n}, \\ \Rightarrow \quad & (ab)^{\frac{n-1}{2}} = a^{\frac{n-1}{2}} b^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right)\left(\frac{b}{n}\right) \equiv \left(\frac{ab}{n}\right) \pmod{n} \\ \Rightarrow \quad & (ab) \in C_n. \end{aligned}$$

Let $A = a^{-1}$ that is $Aa = aA = 1 \mod n$. We show that $A \in C_n$.

$$\left(\frac{a}{n}\right)\left(\frac{A}{n}\right) \equiv \left(\frac{aA}{n}\right) \equiv \left(\frac{1}{n}\right) \equiv 1 \pmod{n} \Rightarrow \left(\frac{a}{n}\right) \equiv \left(\frac{A}{n}\right) \pmod{n} \tag{11.12}$$

Furthermore,

$$\begin{aligned} a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n} \quad & \Rightarrow \quad (Aa)^{\frac{n-1}{2}} \equiv A^{\frac{n-1}{2}} \left(\frac{a}{n}\right) \pmod{n} \\ & \Rightarrow \quad 1 \equiv A^{\frac{n-1}{2}} \left(\frac{a}{n}\right) \pmod{n} \\ & \Rightarrow \quad \left(\frac{a}{n}\right) \equiv A^{\frac{n-1}{2}} \pmod{n} \\ & \xrightarrow{\text{by Eq. (11.12)}} \quad \left(\frac{A}{n}\right) \equiv A^{\frac{n-1}{2}} \pmod{n}, \end{aligned} \tag{11.13}$$

with the latter implying $A \in C_n$. Given that C_n is a subgroup of $\mathbb{U}_n$, the order of C_n divides that of $\mathbb{U}_n$. If the two are not equal then

$$|C_n| \leq \frac{|\mathbb{U}_n|}{2}.$$

We dismiss the possibility $C_n = \mathbb{U}_n$.

(b) Show $C_n \neq \mathbb{U}_n$. The proof is by contradiction (contrapositive). Let us assume that $C_n = \mathbb{U}_n$. Moreover, let $n = p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k}$. We distinguish two cases for n.

Case 1: $a_1 = 1$. For convenience rewrite

$$n = p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k} = p_1 p_2^{a_2} \ldots p_k^{a_k} = p_1 Q,$$

where $\gcd(p_1, Q) = 1$.

Let g be a generator of $\mathbb{Z}_{p_1}^* = \mathbb{U}_{p_1}$. From prior results we know it exists for prime p_1. (Note n is odd.) Moreover $g^{\phi(p_1)} = g^{p_1 - 1} \equiv 1 \pmod{p_1}$ and therefore the following is also true

$$g^{\frac{p_1-1}{2}} \equiv -1 \pmod{p_1},$$

by Lagrange's theorem. Then by the Chinese Remainder Theorem (CRT) there exists an a such that

$$\begin{aligned} a &\equiv g \pmod{p_1}, \\ a &\equiv 1 \pmod{Q}. \end{aligned}$$

We then obtain the following.

$$\begin{aligned}\left(\frac{a}{n}\right) &= \left(\frac{a}{p_1^{a_1}p_2^{a_2}\cdots p_k^{a_k}}\right)\\ &= \left(\frac{a}{p_1 p_2^{a_2}\cdots p_k^{a_k}}\right)\\ &= \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2^{a_2}\cdots p_k^{a_k}}\right)\\ &= \left(\frac{a}{p_1}\right)\left(\frac{a}{Q}\right).\end{aligned}$$

For $\left(\frac{a}{p_1}\right)$ since $a \equiv g \pmod{p_1}$ we have

$$\left(\frac{a}{p_1}\right) \equiv \left(\frac{g}{p_1}\right) \equiv g^{\frac{p_1-1}{2}} \equiv -1 \pmod{p_1}.$$

For $\left(\frac{a}{Q}\right)$ we have by way of $a \equiv 1 \pmod{Q}$ the following

$$\left(\frac{a}{Q}\right) \equiv \left(\frac{1}{Q}\right) \equiv 1 \pmod{Q}.$$

Therefore we conclude the following

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{Q}\right) = (-1)(+1) = -1.$$

If $a \in C_n$ and $\left(\frac{a}{n}\right) = -1$ this implies

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \equiv -1 \pmod{n},$$

Then, we have the following, noting that $n = p_1 Q$ and thus $Q|n$ and $\gcd(p_1, Q) = 1$.

$$\begin{aligned} a^{\frac{n-1}{2}} \equiv -1 \pmod{n} &\Rightarrow n|a^{\frac{n-1}{2}}+1\\ &\Rightarrow Q|a^{\frac{n-1}{2}}+1\\ &\Rightarrow a^{\frac{n-1}{2}} \equiv -1 \pmod{Q} \end{aligned} \tag{11.14}$$

By way of $a \equiv 1 \pmod{Q}$ we also have the following.

$$a^{\frac{n-1}{2}} \equiv 1 \pmod{Q}, \tag{11.15}$$

and therefore combining Eq. 11.14 and Eq. 11.15 we obtain the following

$$2 \equiv 0 \pmod{Q},$$

which is only possible for Q being equal to 2, but this contradicts the fact that $n = p_1 Q$ is an odd integer.
Case 2: $a_1 \geq 2$**.**
If $a_1 \geq 2$ it means $a_1 - 1 \geq 1$. For $C_n = \mathbb{U}_n$, it means

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n} \Rightarrow a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \equiv \pm 1 \pmod{n},$$

and by squaring we obtain the following

$$a^{n-1} \equiv 1 \pmod{n}.$$

This last equation implies $n|a^{n-1}-1$. Since $p_1^{a_1}$ is a factor of n we also have that $p_1^{a_1}|a^{n-1}-1$, or equivalently

$$a^{n-1} \equiv 1 \pmod{p_1^{a_1}}.$$

By prior results the last one implies

$$\phi(p_1^{a_1})|n-1 \Rightarrow p_1^{a_1-1}(p_1-1))|n-1 \xrightarrow{\text{by } a_1>1} p_1|n-1.$$

In addition to $p_1|n-1$ we have $p_1|n$ as p_1 is a factor of n. The two imply $p_1|1$ i.e. $p_1=1$ which contradicts the fact that p_1 is a prime number, and an odd integer and thus $p_1 \geq 3$.
(c) Conclude $C_n \neq \mathbb{U}_n$. By case 1 and case 2 we conclude that $C_n \neq \mathbb{U}_n$. Then

$$|C_n| \leq |\mathbb{U}_n|/2 \leq (n-1)/2$$

Note that for prime N $|\mathbb{U}_n| = \phi(n) = n-1$.
The probability that the output of Algorithm Solovay-Strassen is Composite given that n is a composite number is at least $1-2^{-t}$ as the algorithm repeats this test t times. The result then follows. ■

Conrad [11] establishes a result which is equivalent to this result.

11.4 Probabilistic primality tests

11.4.1 The Miller-Rabin primality test

Exercise 1025.

Show the following lemma. It is used to establish Proposition 11.1 and prove the correctness of Miller-Rabin which is Algorithm 5. A condition expressed by Eq.(11.16) will have cases P_{2a}, P_{2b} merged into Case P_2. Therefore the condition of Eq.(11.16) will appear as in Eq.(11.17) in Lemma 11.1.

$$\text{n is prime} \Rightarrow \forall x, 1 \leq x < n : \begin{cases} x^l \equiv 1 \pmod{n} & \text{Case } P_1(x,n) \\ \vee & \\ x^l \equiv -1 \pmod{n} & \text{Case } P_{2a}(x,n,m) \\ \vee & \\ x^{2^m l} \equiv -1 \pmod{n}, \quad \exists m : 0 < m < k & \text{Case } P_{2b} \end{cases} \tag{11.16}$$

Lemma 11.1

Let $n \in \mathbb{Z}_+$ be an odd integer with $n > 2$. Let $n-1 = 2^k \cdot l$, where $k \geq 1$, and l odd. For all x such that $1 \leq x < n$ we have the following.

$$\text{n is prime} \Rightarrow \forall x, 1 \leq x < n : \begin{cases} x^l \equiv 1 \pmod{n} & \text{Case } P_1(x,n) \\ \vee & \\ x^{2^m l} \equiv -1 \pmod{n}, \quad \exists m : 0 \leq m < k & \text{Case } P_2(x,n,m) \end{cases} \tag{11.17}$$

Solution.

If n is a prime number then Fermat's little theorem is invoked. For any x such that $1 \leq x < n$ and we note that then $\gcd(x,n) = 1$ we have the following.

$$x^{n-1} \equiv 1 \pmod{n} \Rightarrow x^{n-1} - 1 \equiv 0 \pmod{n}.$$

We note that $n-1 = 2^k l$ as stated therefore

$$n-1 = \frac{n-1}{2} \cdot 2 = 2^{k-1} l + 2^{k-1} l = 2 \cdot (2^{k-1} l).$$

Therefore we have the following

$$x^{n-1} - 1 = x^{2^k l} - 1 = x^{2 \cdot 2^{k-1} l} - 1 \quad = \quad \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-1} l} - 1\right) \tag{11.18}$$

We continue unrolling the right-most term a total of k times.

$$\begin{aligned} x^{n-1} - 1 &= \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-1} l} - 1\right) \\ &= \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-2} l} + 1\right)\left(x^{2^{k-2} l} - 1\right) \\ &= \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-2} l} + 1\right)\left(x^{2^{k-3} l} + 1\right)\left(x^{2^{k-3} l} - 1\right) \\ &\cdots \\ &= \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-2} l} + 1\right) \ldots \left(x^{2l} + 1\right)\left(x^{l} + 1\right)\left(x^{l} - 1\right). \end{aligned} \tag{11.19}$$

Moreover, $x^{n-1} - 1 \equiv 0 \pmod{n}$ implies the following.

$$\begin{aligned} x^{n-1} - 1 &\equiv 0 \pmod{n} \Leftrightarrow \\ \left(x^{2^{k-1} l} + 1\right)\left(x^{2^{k-2} l} + 1\right) \ldots \left(x^{2l} + 1\right)\left(x^{l} + 1\right)\left(x^{l} - 1\right) &\equiv 0 \pmod{n} \end{aligned} \tag{11.20}$$

We perform a case analysis.
One possibility that generates P_1 is that $n|x^l-1$, the right-most term. Then $x^l \equiv 1 \pmod{n}$. If this is not the case, then n might divide the second term from the right that is, $n|x^l+1$. Then $x^l \equiv -1 \pmod{n}$. This is what was called case P_{2a} or it is the first case of P_2 for $m=0$. If this is not the case, then n might divide the third term from the right that is, $n|x^{2l}+1$. Then $x^{2l} \equiv -1 \pmod{n}$. This is the first case of P_{2b} with $m=1$. Continuing like this, if it is not the case that n divides $x^{2^{k-2}l}+1$, then $n|x^{2^{k-1}l}+1$. Then $x^{2^{k-1}l} \equiv -1 \pmod{n}$. This is the last case of case P_{2b} with $m=k-1$.

Note that $x^{2^{k-1}l}$ is the square of $x^{2^{k-2}l}$ and so on, and x^{2l} is the square of $x^l \bmod n$. Thus the sequence implied is as follows if listed right to left.

$$x^l \bmod n, x^{2l} \bmod n, x^{2^2 l} \bmod n, \ldots x^{2^{k-1}l} \bmod n \tag{11.21}$$

The terms as shown in this order can be generated by squaring mod n starting with the left-most term $x^l \bmod n$. If x^l is equivalent to one mod n so are all subsequent terms. If $x^{2^m l}$ is equivalent to 1 mod n so are all subsequent terms $x^{2^j l}$, $m \le j < k$ because of squaring. If $x^{2^m l}$ is -1 mod n all subsequent terms $x^{2^j l}$, $m < j < k$, are 1 mod n. For the sequence This implies that in the sequence of Eq. (11.21) there is only one -1 followed by (a suffix) of ones to the end of it or the whole sequence is a sequence of ones.
In conclusion if n is prime then we have P_1 or P_2 that is

$$\text{n is prime} \Rightarrow \forall x, 1 \le x < n : P_1(x,n) \vee \exists 0 \le m < k : P_2(x,n,m). \tag{11.22}$$

If n is a composite number then we have $\overline{P_1 \vee P_2} = \overline{P_1} \wedge \overline{P_2}$.

$$\text{n is composite} \Rightarrow \exists x, 1 \le x < n : \overline{P_1(x,n)} \wedge \forall m, 0 \le m < k : \overline{P_2(x,n,m)}. \tag{11.23}$$

This is equivalent to the following one.

$$\text{n is composite} \Leftrightarrow \exists x, 1 \le x < n : \begin{cases} x^l \not\equiv 1 \pmod{n} & \text{Case } \overline{P_1(x,n)} \\ \wedge & \\ x^{2^m l} \not\equiv -1 \pmod{n}, \quad \forall m, 0 \le m < k : & \text{Case } \overline{P_2(x,n,m)} \end{cases} \tag{11.24}$$

Later one we will rename

$$C_1(x,n) = \overline{P_1(x,n)} \tag{11.25}$$
$$C_2(x,n,m) = \overline{P_2(x,n,m)}. \tag{11.26}$$

■

Exercise 1026.

We refine Lemma 11.1 to match the corresponding lines of Algorithm 5. Show the following proposition.

Proposition 11.1

Let $n \in \mathbb{Z}_+$ be an odd integer with $n > 4$. Let $n-1 = 2^k \cdot l$, where $k \geq 1$, and l odd. Show then, integer n is declared a PseudoPrime number if and only if for every a such that $2 \leq a < n-1$ and $n \nmid a$ either

$$P_1(a,n) = \overline{C_1(a,n)} \quad : \quad a^l \equiv 1 \pmod{n}, \tag{11.27}$$

or

$$P_2(a,n,m) = \overline{C_2(a,n,m)} \quad : \quad a^{2^m l} \equiv -1 \pmod{n}, \tag{11.28}$$

for some integer m such that $0 \leq m < k$. PseudoPrime means it is either a prime number or $a^{n-1} \equiv 1 \pmod{n}$.

We note that the conditions of Proposition 11.1 are Eq. (11.22) of Lemma 11.1 for $x = a$. Therefore the proof of correctness directly follows from Lemma 11.1. The proof of correctness of Proposition 11.1 deals with the necessary and sufficient part, whereas Lemma 11.1 did no elaborate on the latter.

Solution.
$\Rightarrow$ (Sufficient condition.)
This is Lemma 11.1 summarized. Let n be a prime number. Then by Fermat's little theorem we have the following

$$a^{n-1} \equiv 1 \pmod{n},$$

for all $a \in \mathbb{Z}$ such that $n \nmid a$. Thefore we have the following.

$$\begin{aligned} a^{n-1} &\equiv 1 \pmod{n} \Leftrightarrow \\ a^{2^k l} &\equiv 1 \pmod{n} \Leftrightarrow \\ (a^{2^{k-1} l} - 1)(a^{2^{k-1} l} + 1) &\equiv 0 \pmod{n} \Rightarrow \\ a^{2^{k-1} l} &\equiv \pm 1 \pmod{n}. \end{aligned}$$

The last modular equation gives one of the following. $n | a^{2^{k-1} l} - 1$ or $n | a^{2^{k-1} l} + 1$, for a prime number p. We distinguish and examine the two cases separately.
Case 1. If $a^{2^{k-1} l} \equiv -1 \pmod{n}$, we are done. This is the second case in the statement.
Case 2. If $a^{2^{k-1} l} \not\equiv -1 \pmod{n}$, then $a^{2^{k-1} l} \equiv 1 \pmod{n}$, and we repeat the same argument to conclude $n | a^{2^{k-2} l} - 1$ or $n | a^{2^{k-2} l} + 1$.
Likewise, at some point $a^{2^m l} \equiv -1 \pmod{p}$, for some m such that $0 \leq m < k$, and we stop by way of the second case of the statement again, or we exhaust the 2^m to reach $a^l \equiv -1 \pmod{n}$, or $a^l \equiv 1 \pmod{n}$. In the former case, we reach the second case of the statement for $m = 0$; in the latter case we reach the first case of the statement. In either case we are done. Sufficient condition proved.
$\Leftarrow$ (Necessary condition.)
We distinguish two cases.
Case 1. Let $a^l \equiv 1 \pmod{n}$. Then we have the following.

$$\begin{aligned} a^l &\equiv 1 \pmod{n} \Rightarrow \\ a^{2^k l} &\equiv 1 \pmod{n} \Rightarrow \\ a^{n-1} &\equiv 1 \pmod{n}. \end{aligned}$$

Case 2. Let $a^{2^m l} \equiv -1 \pmod{n}$. Then we also observe the following. $a^{2^m l} \not\equiv 1 \pmod{n}$, since otherwise $n = 2$,

and n is trivially prime. We then have the following.

$$\begin{aligned} a^{2^m l} &\equiv -1 \pmod{n} \Leftrightarrow \\ a^{n-1} = \left(a^{2^m l}\right)^{2^{k-m}} = (-1)^{2^{k-m}} &\equiv 1 \pmod{n}, \end{aligned}$$

since $m < k$ and thus $k - m > 0$ and $2^{k-m} \geq 2$. ■

Exercise 1027.

Theorem 11.3

Algorithm 5 is a probabilistic algorithm for testing whether n is composite or not.
Prove the correctness of Algorithm 5.

Solution.
Let $P_1(a,n)$ and $P_2(a,n,m)$ be the two conditions of Proposition 11.1 as indicated in Eq. (11.22) of Lemma 11.1 for $x=a$. Let $C_1(a,n)$ and $C_2(a,n,m)$ be the two conditions of Eq.(11.25) and Eq.(11.26) established respectively from Eq. (11.24) of Lemma 11.1 for $x=a$. They are also in negation form part of Eq.(11.27) and Eq.(11.28) in Proposition 11.1.
In line 2 of Algorithm 4, or line 4 of Algorithm 5, an a is picked uniformly at random such that $2 \leq a < n-1$.
Step 1. The condition $n \nmid a$ of Proposition 11.1 translates into $\gcd(a,n) > 1$ since $a < n$, and this maps to lines 3-5 of Algorithm 4, or lines 5-7 of Algorithm 5. If n is composite the algorithm terminates with the correct answer Composite and a becomes a witness of the compositeness of n.
We are left with a sequence of steps to determine other wise the compositeness of n for an a such that $\gcd(a,n) = 1$.
Step 2. If

$$C_1(a,n) \equiv \overline{P_1(a,n)}: \quad a^l \not\equiv 1 \pmod{n},$$

and

$$C_2(a,n,m) \equiv \overline{P_2(a,n,m)}: a^{2^m l} \not\equiv -1 \pmod{n},\ \forall m, 0 \leq m < k,$$

report n to be composite and a to be the witness of its compositeness. This is a consequence of Eq.(11.24) of Lemma 11.1.
Step 3. Otherwise n is either prime or composite.
A single execution of Steps 1, 2, and 3 is Algorithm 4. Repeat Step 1,2, and 3 a total of t times and this leads to Algorithm 5.
Algorithm 5 tests for the two conditions $C_1(a,n)$ and $C_2(a,n,m)$ a number of times. We denote the number of times these conditions are tested as t. Every time these conditions are tested we pick a random a from the interval/set $\{2,3,\ldots,n-2\}$ of cardinality $n-1$. If for the chosen a, we have $\overline{P_1(a,n)}$ and $\overline{P_2(a,n,m)}$ for all applicable m, then n is composite and a is the Rabin-Miller witness of the compositeness of n and denoted as MRw(a,n), or the set of witnesses for n is denoted as MRw(n). Otherwise a is a non-witness denoted MRnw(a,n) and the corresponding set MRnw(n). To simplify things we used $C_1(a,n) = \overline{P_1(a,n)}$ and $C_2(a,n,m) = \overline{P_2(a,n,m)}$.

$$MRw(a,n) \equiv a \in MRw(n) \Leftrightarrow C_1(a,n) \wedge \left(\wedge_{m=0}^{k-1} C_2(a,n,m)\right).$$

We obtain the following condition for non witnesses MRnw(a,n) or set MRnw(n).

$$MRnw(a,n) \equiv a \in MRnw(n) \Leftrightarrow \overline{C_1(a,n)} \vee \left(\vee_{m=0}^{k-1} \overline{C_2(a,n,m)}\right) \Leftrightarrow P_1(a,n) \vee \left(\vee_{m=0}^{k-1} P_2(a,n,m)\right).$$

Line 18 of Algorithm 4 is equivalent to a return of PseudoPrime; whether it is done through line 18 or line 24 is the difference of executing the conditional of line 21.

There are two conditions that can be checked together: these are $C_1(a,n)$ and $C_2(a,n,0)$ that map to $a^l \not\equiv 1 \pmod{n}$ and to $a^l \not\equiv -1 \pmod{n}$ respectively. The former is checked in line 9 of Algorithm 5 and the latter in line 15, $m=0$ of Algorithm 5 The condition $C_2(a,n,m)$ for all m such that $0 \leq m < k$ is checked in lines 15-22 of Algorithm 5. It includes the check $C_2(a,n,0)$. Similarly for Algorithm 4.

If $a^l \not\equiv 1 \pmod{n}$ is not true i.e. $C_1(a,n)$ fails, there is no reason to continue with $C_2(a,n,m)$ checking. Algorithm 4 exits through line 10. Algorithm 5 continues its execution. If $a^l \not\equiv 1 \pmod{n}$ is true i.e. $C_1(a,n)$ succeeds we continue with $C_2(a,n,m)$. For a composite n, $C_2(a,n,m)$ must check true for all $m = 0,\ldots,k-1$. Thus if Condition is true, this is the case and Algorithm 5 exits through line 24, otherwise an $m=j$ that results in a failure of $C_2(a,n,j)$, forces an exit through line 20 of the for loop for lines 15-22; in that case Condition is false,

Algorithm 4: MillerRabinCTest (n): Miller-Rabin compositeness test

Input: $n > 4$ is odd $n-1 = 2^k l$, l odd
Output: n is **Composite** or **Pseudoprime**

$n-1 = 2^k \cdot l$;
Pick $a \in \{2,3,\ldots,n-2\}$ uniformly at random ;
if $\gcd(a,n) > 1$ **then**
  return Composite ; /* a is a witness of n's compositeness */
end
$x = a^l \pmod{n}$;
if $a^l \not\equiv 1 \pmod{n}$ **then** /* $C_1(a)$ */
  ;
else
  return PseudoPrime
end
Condition = TRUE ;
for $m = 0,1,\ldots,k-1$ **do** /* $C_2(a,n,m)$ */
  if $a^{2^m l} \not\equiv -1$ **then**
    Condition = Condition $\wedge$ TRUE;
  else
    Condition = Condition $\wedge$ FALSE;
    break ; /* or return(**PseudoPrime**) */
  end
end
if *Condition* **then**
  return Composite ; /* a is a witness of n's compositeness */
else
  return PseudoPrime
end

and the algorithm continues until the limit of t iterations is reached. Therefore, if the algorithm is unsuccessful in proving the compositeness of n, Algorithm 5 exits through line 29. Pseudoprime then means that n can be composite (no proof of the compositeness of n was found) or it can ble prime (and this is the reason no proof of the compositeness of n was found).

Algorithm 5: Miller-Rabin primality test algorithm

Input: $n > 4$ is odd $n-1 = 2^k l$, l odd; number of runs is t
Output: n is **Composite** or **Pseudoprime**

```
1  n-1 = 2^k l ; i = 0 ;
2  do
3  |  i = i+1 ;
4  |  Pick a ∈ {2,3,...,n-2} uniformly at random ;
5  |  if gcd(a,n) > 1 then
6  |  |  return Composite ;                    /* a is a witness of n's compositeness */
7  |  end
8  |  x = a^l (mod n);
9  |  if a^l ≢ 1 (mod n) then                                   /* C1(a,n) */
10 |  |  ;
11 |  else
12 |  |  continue;
13 |  end
14 |  Condition = TRUE ;
15 |  for m = 0,1,...,k-1 do                                    /* C2(a,n,m) */
16 |  |  if a^(2^m l) ≢ -1 then                  /* x = x^2 mod n; x ≢ -1 (mod n) */
17 |  |  |  Condition = Condition ∧ TRUE;
18 |  |  else
19 |  |  |  Condition = Condition ∧ FALSE;
20 |  |  |  break;
21 |  |  end
22 |  end
23 |  if Condition then
24 |  |  return Composite
25 |  else
26 |  |  continue ;
27 |  end
28 while i < t;
29 return PseudoPrime ;                          /* n is either prime or composite */
```

Moreover, note that the powers of line 16 are generated from a^l i.e. the statement of line 8 through repeated doubling as indicated in the comment of line 16. We do not need to generate powers from scratch (starting from a) every time. Powers are mod n always. Therefore a -1 mod n can be rewritten $n-1$ mod n. ■

Exercise 1028.
Show the proposition below.

Proposition 11.2 **Probability of successful reporting**

If n is an odd composite number greater than four then Algorithm 5 returns Composite with probability at least $3/4$. The probability that the output of Algorithm 5 is Composite given that n is a composite number is at least $1-4^{-t}$.

Solution. It was proved by Rabin [32] and independently by Monier [28] that for a composite odd n, $|MRw(a)|$ is at least $3\phi(n)/4$ and $|MRnw(a)|$ is thus at most $\phi(n)/4$. But the result is applicable to Miller1. Algorithm 5 uses the conditions of Miller2. Monier [28] proved that the conditions used in Miller1 and Miller2 are equivalent. So we can use for this algorithm the claims of Rabin [32] and also of Monier [28]. Later, we introduce RabinMiller which was analyzed by Rabin [32]. ■

Exercise 1029.

Algorithm 6 does not use the conditions of Eq.(11.27) and Eq.(11.28), where

$$P_1(a,n) = \overline{C_1(a,n)} \quad : \quad a^l \equiv 1 \pmod{n}, \tag{11.29}$$

$$P_2(a,n,m) = \overline{C_2(a,n,m)} \quad : \quad a^{2^m l} \equiv -1 \pmod{n}, \tag{11.30}$$

It uses instead the following conditions of Eq.(11.34) and Eq.(11.33), where

$$R_1(a,n) = \overline{M_1(a,n)} \quad : \quad a^{n-1} \not\equiv 1 \pmod{n}, \tag{11.31}$$

$$R_2(a,n,m) = \overline{M_2(a,n,m)} \quad : \quad d_m \neq 1 \wedge d_n \neq n. \tag{11.32}$$

where $d_m = \gcd(a^{2^m \cdot l} - 1, n)$, $m \in \mathbb{N}$.

Then instead of using Eq.(11.17) and Eq.(11.24) it establishes a new set of conditions. We first rewrite Eq.(11.17) and Eq.(11.24), the conditions used by Miller 2.

$$\text{n is prime} \Rightarrow \forall a, 1 \leq a < n : \begin{cases} a^l \equiv 1 \pmod{n} & \text{Case } P_1(a,n) \\ \vee & \\ a^{2^m l} \equiv -1 \pmod{n}, \quad \exists m : 0 \leq m < k & \text{Case } P_2(a,n,m) \end{cases} \tag{11.17}$$

$$\text{n is composite} \Leftrightarrow \exists a, 1 \leq a < n : \begin{cases} a^l \not\equiv 1 \pmod{n} & \text{Case } \overline{P_1(a,n)} \\ \wedge & \\ a^{2^m l} \not\equiv -1 \pmod{n}, \quad \forall m, 0 \leq m < k : & \text{Case } \overline{P_2(a,n,m)} \end{cases} \tag{11.24}$$

Then we write the new set of conditions used by Miller1.

$$\text{n is prime} \Rightarrow \forall a : \quad n \nmid a \wedge \begin{cases} a^{n-1} \equiv 1 \pmod{n} & \text{Case } M_1(a,n) = \overline{R_1(a,n)} \\ \wedge & \\ \forall m, 0 \leq m < k : d_m = 1 \vee d_m = n & \text{Case } M_2(a,n,m) = \overline{R_2(a,n,m)} \end{cases} \tag{11.33}$$

$$\text{n is composite} \Leftrightarrow \exists a : \quad n|a \quad \vee \quad \begin{cases} a^{n-1} \not\equiv 1 \pmod{n} & \text{Case } R_1(a,n) \\ \vee & \\ \exists m, 0 \leq m < k : d_m \neq 1 \wedge d_m \neq n & \text{Case } R_2(a,n,m) \end{cases} \tag{11.34}$$

Show the following.

Proposition 11.3 **Equivalence of condition in Miller1 and Miller 2**

Let n be an odd number $n > 4$ and that $n - 1 = 2^k \cdot l$, $k \geq 1$ and l is odd. The conditions of Eq.(11.33) and Eq.(11.34) utilized by MillerRabin are equivalent to the conditions of Eq.(11.17) and Eq.(11.24) utilized by RabinMiller.

Solution.

For both set of conditions we prove the equivalence for n being a prime number. The case n is composite can be proved by symmetry using De Morgan's Law. Consider the sequence of the d values forming the following row vector in that case.

$$(d_0, d_1, \ldots, d_{k-1}, d_k).$$

We will show first that if a d_j value is n then all the values to its right will also be n that is $d_i = n$ for all $i > j$. Thus 1s can extend only to the left of the leftmost n. As a conclusion the row vector can be: (a) all n, or (b) a

sequence of consecutive 1s followed by consecutive ns, and (c) the rightmost position is n.
Condition $M_1(a,n)$. By way of $M_1(a,n)$ we have the following

$$\begin{aligned} a^{n-1} \equiv 1 \pmod{n} &\Rightarrow n|a^{n-1}-1 \Rightarrow n|a^{2^k l}-1 \\ &\Rightarrow a^{2^k l}-1 \equiv 0 \pmod{n} \Rightarrow \gcd(a^{2^k l}-1,n)=n \\ &\Rightarrow M_a(a,k): d_k = n. \end{aligned} \tag{11.35}$$

Therefore we obtain that the right-most position of the row vector has $d_k = n$.

$$(d_0, d_1, \ldots, d_{k-1}, n = d_k).$$

Condition $M_2(a,n,m)$ case (i): for some $0 \leq j < k$: $d_j = n$. We prove that a n in some position j is followed to the right with a block of ns. Consider that for some m, call it j, we have for $M_2(a,n,j)$ the following: $d_j = n$. This implies $d_j = n = \gcd(a^{2^j l}-1,n)$, and also $a^{2^j l} \equiv 1 \pmod{n}$. Since $a^{2^{j+1} l} = (a^{2^j l})^2$, $a^{2^{j+1} l} \equiv 1 \pmod{n}$, and therefore $\gcd(a^{2^{j+1} l}-1,n) = d_{j+1} = n$, thus having $d_{j+1} = n$. By induction, all $i > j$ have $d_i = n$ as well. Therefore we obtain the following refinement for case (i).

$$(d_0, d_1, \ldots, d_{j-1}, d_j = n, d_{j+1} = n, \ldots, d_{k-1} = n, d_k = n).$$

If $j = 0$ we have an all n row-vector.

$$(d_0 = n, d_1 = n, \ldots, d_{k-1} = n, d_k = n).$$

We conclude that $d_j = n$ is equivalent to $a^{2^j l} \equiv 1 \pmod{n}$.
Condition $M_2(a,n,m)$ case (ii): if the left-most n is at position $j > 0$ i.e. $d_j = n$ there is a $d_{j-1} = 1$. Consider that for some m, call it j, we have for $M_2(a,n,j)$ the following: $d_j = n$ and it is the leftmost n entry of the row vector. The latter implies $\gcd(a^{2^j l}-1,n) = d_j = n$ or $a^{2^j l} \equiv 1 \pmod{n}$. Let us consider $m = j-1$. We have $d_{j-1} = 1$.
Let us set $x = a^{2^{j-1} l}$. Then $x^2 = a^{2^j l}$. Since $a^{2^j l} \equiv 1 \pmod{n}$, we have $x^2 \equiv 1 \pmod{n}$.

$$\begin{aligned} x^2 &\equiv 1 \pmod{n} \\ (x-1)(x+1) &\equiv 0 \pmod{n} \\ (x-1)(x+1) &\equiv 0 \pmod{p_i^{a_i}} \\ (x+1) &\equiv 0 \pmod{p_i^{a_i}} \\ x &\equiv -1 \pmod{p_i^{a_i}}, \end{aligned}$$

and given $\gcd(p_u^{a_u}, p_v^{a_v}) = 1$, we have $x \equiv -1 \pmod{n}$. We need to explain the step that led to the dismisal of $(x-1)$. We claimed that $\gcd(p_i^{a_i}, x-1) = 1$. This is because of the following.

$$\begin{aligned} 1 = d_{j-1} &= \gcd(a^{2^{j-1} l}-1,n) \\ &= \gcd(x-1,n) \\ &= \gcd(x-1, p_i^{a_i}) \forall i \end{aligned}$$

In conclusion $a^{2^{j-1} l} \equiv -1 \pmod{n}$. On the left of a 1 we can only have 1. If there was an n then to its right there should have been an n and not an 1. We conclude that $d_j = 1$ is equivalent to $a^{2^j l} \equiv -1 \pmod{n}$. Now that we examined the cases for $M_1(a,n)$ and $M_2(a,n,m)$, we move to $P_1(a,n)$ and $P_2(a,n,m)$.
Mapping M_1, M_2 to P_1, P_2.
A row vector that has n in the left-most position means $d_0 = n$ which is equivalent to $a^{2^0 l} \equiv a^l \equiv 1 \pmod{n}$.

This is case $P_1(a,n)$. A row vector that has 1 in the left-most position $m=0$ and n in position $m=1$, by prior discussion (case (ii)) $a^l \equiv -1 \pmod{n}$ which is $P_2(a,n,0)$. If the right-most 1 is in position $j-1$ then $a^{2^{j-1}l} \equiv -1 \pmod{n}$ which is $P_2(a,n,j-1)$. The rightmost m for a $d_j = 1$ is $j = k-1$. Since $d_k = n$ this maps to $a^{2^{k-1}l} \equiv -1 \pmod{n}$ which is $P_2(a,n,k-1)$. There is a 1-1 map of $M_1(a,n) \mapsto P_1(a,n)$ and $M_2(a,n,m) \mapsto P_2(a,n,m)$. This concludes the proof.

Supplement 1: Show $a^{2^{j+1}l} \equiv 1 \pmod{n}$ if $a^{2^j l} \equiv 1 \pmod{n}$. Let $n = p_1^{a_1} \dots p_r^{a_r}$ be the prime decomposition of n. $a^{2^j l} \equiv 1 \pmod{n}$ implies for every $i = 1, \dots, r$ that $a^{2^j l} \equiv 1 \pmod{p_i^{a_i}}$.

$$\begin{aligned} a^{2^j l} &\equiv 1 \pmod{p_i^{a_i}} \\ (a^{2^j l} - 1)^2 &\equiv 0 \pmod{p_i^{a_i}} \\ a^{2^{j+1}l} + 1 - 2a^{2^{j+1}} l &\equiv 0 \pmod{p_i^{a_i}} \\ a^{2^{j+1}l} + 1 - 2 &\equiv 0 \pmod{p_i^{a_i}} \\ a^{2^{j+1}l} &\equiv 1 \pmod{p_i^{a_i}}, \end{aligned}$$

and given $\gcd(p_u^{a_u}, p_v^{a_v}) = 1$, we have $a^{2^{j+1}l} \equiv 1 \pmod{n}$ i.e. $d_{j+1} = n$.

■

Exercise 1030.

Rabin [32] converted Algorithm 6 into a probabilistic primality test algorithm. Show the following proposition.

Algorithm 6: Miller1(n): primality test algorithm

Input: $n > 4$ is odd; $n-1 = 2^k l$, l odd
Output: n is **Composite** or **Prime**

```
1  n-1 = 2^k l ; f(n) = c(ln n)^2 ;                    /* Bach [3] calculated c = 2 */
2  if PerfectPower(n) then                                              /* Line 1 */
3  |  return Composite ;                          /* n is a perfect power n = p^s, s ≥ 2 */
4  end
5  for a = 2,3,...,2(ln n)^2 do
6  |  if (a|n)||(gcd(a,n) > 1) then                                  /* Line 2 (i) */
7  |  |  return Composite ;                                  /* Just a|n in [27] */
8  |  end
9  |  if a^(n-1) ≢ 1 (mod n) then                                   /* Line 2 (ii) */
10 |  |  return Composite ;
11 |  end
12 |  if ∃m, 0 ≤ m < k : gcd((a^(2^k l) mod n) - 1, n) ≠ 1, n then  /* Line 2 (iii) */
13 |  |  return Composite ;
14 |  end
15 end
16 return Prime ;                                                      /* Line 3 */
```

Proposition 11.4 **Probability of successful reporting**

Algorithm 7 is a probabilistic algorithm for testing whether n is prime or composite. If n is an odd composite number greater than four then Algorithm 7 returns Composite with probability at least $3/4$. The probability that the output of Algorithm 7 is Composite given that n is a composite number is at least $1-4^{-t}$.

Solution.
Let $W(a,n)$ [32] indicate that a is a witness of the compositeness of n. Then denote by $W(n)$ the set of all $W(a,n)$ witnesses. Later on we we will use the terminology $MRw(a,n)$ and $MRw(n)$. for $W(a,n)$ and $W(n)$ respectively.

$$W(n) = \{a : W(a,n)\}.$$

It was proved in [32] that for $n > 4$ and n composite

$$|W(n)| \geq \frac{3(n-1)}{4}$$

In a theorem to follow, it will be shown that

$$|MRnw(n)| \leq \frac{\phi(n)}{4} \leq \frac{n-1}{4},$$

that implies

$$|MRw(n)| \geq \frac{3(n-1)}{4}.$$

The proof of correctness appears in [27] and the derivations above in [32]. We present later a proof that is based on [14], with the latter originating from the paper of [28]. It was proved by Rabin [32] and independently by Monier [28] that for a composite odd n, $|MRw(a)|$ is at least $3\phi(n)/4$ and $|MRnw(a)|$ is thus at most $\phi(n)/4$.

Algorithm 7: Rabin-Miller primality test algorithm

Input: $n > 4$ is odd $n-1 = 2^k l$, l odd; number of runs is t
Output: n is **Composite** or **PseudoPrime**

$n-1 = 2^k l$; $i = 0$;
if *PerfectPower(n)* **then** /* Line 1 */
  return Composite ; /* [32] is unclear about this test */
end
do
  $i = i+1$;
  Pick $a \in \{2,3,\ldots,n-2\}$ uniformly at random ;
  if $(a|n)||(\gcd(a,n) > 1)$ **then** /* Line 2 (i) */
    return Composite ; /* [32] is unclear about this test */
  end
  $x = a^l \pmod{n}$;
  Calculate using repeated doubling $a^{2^i l} \bmod n$, for $i = 1,\ldots,k$; /* $a^{2^k l} \bmod n$ is $a^{n-1} \bmod n$ */
  if $a^{n-1} \not\equiv 1 \pmod{n}$ **then** /* Line 2 (ii) */
    return Composite ; /* a is a $W(a,n)$ */
  end
  if $\exists m, 0 \le m < k : \gcd((a^{2^k l} \bmod n) - 1, n) \ne 1, n$ **then** /* Line 2 (iii) */
    return Composite ; /* a is a $W(a,n)$ */
  end
while $i < t$;
return PseudoPrime ; /* Line 3 */

■

Exercise 1031.

Consider $n = 9$, a composite and odd integer $n > 4$. It is $n-1 = 2^k \cdot l = 2^3 \cdot 1$, with l being an odd integer. Determine the MRw(n) and MRnw(n).

Solution. We note that $n = 9$, we have $n-1 = 2^k \cdot l$, where $k = 3$ and $l = 1$. The first column is a. The second column shows $a^l \bmod n$. The third column indicates whether $C_1(a,n)$ is the case i.e. $a^l \not\equiv 1 \pmod{n}$. A $+$ indicates so, a $-$ indicates the opposite i.e. $a^l \equiv 1 \pmod{n}$. The fourth column indicates whether $C_2(a,n,0)$ is the case i.e. $a^l \not\equiv -1 \pmod{n}$. A $+$ indicates so, a $-$ indicates the opposite i.e. $a^l \equiv -1 \pmod{n}$.
The fifth column shows $a^{2^1 \cdot l} \bmod n$. The sixth column indicates with a $+$ whether $C_2(a,n,1)$ is the case i.e. $a^{2^1 \cdot l} \not\equiv -1 \pmod{n}$; otherwise a $-$ is shown. The seventh column shows $a^{2^2 \cdot l} \bmod n$. The eighth column indicates with a $+$ whether $C_2(a,n,2)$ is the case i.e. $a^{2^2 \cdot l} \not\equiv -1 \pmod{n}$; otherwise a $-$ is shown. In the ninth column a $+$ indicates that a is a witness, and an $-$ that a is a non-witness to the compositeness of n.
For $a \in MRw(n)$ we need only $+$'s for the corresponding row. For $a \in MRnw(n)$ we need at least one $-$ or equivalently an n in the MRw(n) column.

a	a^l :	$C_1(a,n)$	$C_2(a,n,0)$	$a^{2^1 \cdot l}$:	$C_2(a,n,1)$	$a^{2^2 \cdot l}$:	$C_2(a,n,2)$	$MRw(n)$
1	1	$-$	$+$	1	$+$	1	$+$	n
2	2	$+$	$+$	4	$+$	7	$+$	y
3	3	$+$	$+$	0	$+$	0	$+$	y
4	4	$+$	$+$	7	$+$	4	$+$	y
5	5	$+$	$+$	7	$+$	4	$+$	y
6	6	$+$	$+$	0	$+$	0	$+$	y
7	7	$+$	$+$	4	$+$	7	$+$	y
8	8	$+$	$-$	1	$+$	1	$+$	n

We observe the number of MRw(a,n) is $\geq \frac{3(n-1)}{4} = 6$. The number of MRnw(a,n) is $\leq \frac{(n-1)}{4} = 2$. ■

Exercise 1032.

Consider $n = 13$, a non-composite (!!) and odd integer $n > 2$. It is $n-1 = 2^2 \cdot 3$. Using the prior definition we have $k = 2$ and $l = 3$, with l being a odd number. Determine the MRw(n) and MRnw(n).

Solution.

We note that $n = 13$, we have $n-1 = 2^2 \cdot 3$, where $k = 2$ and $l = 3$. The first column is a. The second column shows $a^l \bmod n$. The third column indicates whether $C_1(a,n)$ is the case i.e. $a^l \not\equiv 1 \pmod{n}$. A $+$ indicates so, a $-$ indicates the opposite i.e. $a^l \equiv 1 \pmod{n}$. The fourth column indicates whether $C_2(a,n,0)$ is the case i.e. $a^l \not\equiv -1 \pmod{n}$. A $+$ indicates so, a $-$ indicates the opposite i.e. $a^l \equiv -1 \pmod{n}$.

The fifth column shows $a^{2^1 \cdot l} \bmod n$. The sixth column indicates with a $+$ whether $C_2(a,n,1)$ is the case i.e. $a^{2^1 \cdot l} \not\equiv -1 \pmod{n}$; otherwise a $-$ is shown. In the seventh column a $+$ indicates that a is a witness, and an $-$ that a is a non-witness to the compositeness of n.

For $a \in MRw(n)$ we need only $+$'s for the corresponding row. For $a \in MRnw(n)$ we need at least one $-$ or equivalently an n in the MRw(n) column. There are no witnesses to the compositeness of 13; we know that 13 is a prime number not a composite number! We observe the number of MRw(a,n) is 0. The number of MRnw(a,n) is thus $n-1 = 12$.

a	a^l :	$C_1(a,n)$	$C_2(a,n,0)$	$a^{2^1 \cdot 1}$:	$C_2(a,n,1)$	$MRw(n)$
1	1	$-$	$+$	1	$+$	n
2	8	$+$	$+$	12	$-$	n
3	1	$-$	$+$	1	$+$	n
4	12	$+$	$-$	1	$+$	n
5	8	$+$	$+$	12	$-$	n
6	8	$+$	$+$	12	$-$	n
7	5	$+$	$+$	12	$-$	n
8	5	$+$	$+$	12	$-$	n
9	1	$-$	$+$	1	$+$	n
10	12	$+$	$-$	1	$+$	n
11	5	$+$	$+$	12	$-$	n
12	12	$+$	$-$	1	$+$	n

■

Exercise 1033.
Consider $n = 21$, a composite and odd integer $n > 2$. It is $n - 1 = 2^2 \cdot 5$. Using the prior definition we have $k = 2$ and $l = 5$, with l being a odd number. Determine the MRw(n) and MRnw(n).

Solution.

a	a^l :	$C_1(a,n)$	$C_2(a,n,0)$	$a^{2^1 \cdot 1}$:	$C_2(a,n,1)$	$MRw(n)$
1	1	$-$	$+$	1	$+$	n
2	11	$+$	$+$	16	$+$	y
3	12	$+$	$+$	18	$+$	y
4	16	$+$	$+$	4	$+$	y
5	17	$+$	$+$	16	$+$	y
6	6	$+$	$+$	15	$+$	y
7	7	$+$	$+$	7	$+$	y
8	8	$+$	$+$	12	$+$	y
9	18	$+$	$+$	9	$+$	y
10	19	$+$	$+$	4	$+$	y
11	2	$+$	$+$	4	$+$	y
12	3	$+$	$+$	9	$+$	y
13	13	$+$	$+$	1	$+$	y
14	15	$+$	$+$	7	$+$	y
15	14	$+$	$+$	15	$+$	y
16	4	$+$	$+$	16	$+$	y
17	5	$+$	$+$	4	$+$	y
18	9	$+$	$+$	18	$+$	y
19	10	$+$	$+$	16	$+$	y
20	20	$+$	$+$	1	$+$	n

We observe the number of MRw(n) is 18. The number of MRnw(n) is 2. ■

Exercise 1034.

We present a proof of Proposition 11.2 that is based on [14] originating from the paper of [28]. Proposition 11.2 will utilize the result of Theorem 11.4. Moreover, Theorem 11.4 relies on a series of propositions itself.

> **Theorem 11.4** **MillerRabin bound**
>
> *In a follow-up problem the following will be shown.*
>
> $$|C(n)| \leq \phi(n)/4,$$
>
> *which would then imply*
>
> $$|MRw(n)| \geq (n-1) - \phi(n)/4.$$

Show this theorem after showing the four propositions quoted in its proof.

Solution. It follows by way of Proposition 11.5 shown earlier, Proposition 11.6, Proposition 11.7, and Proposition 11.8 below. ■

Exercise 1035.
Show the following.

Proposition 11.5

Let $n \in \mathbb{Z}_+$ be an odd composite integer with $n > 10$, such that $n - 1 = 2^k l$, where l is an odd integer. Furthermore let $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ be the prime decomposition of n, and let $d(n) = r$ indicate the number of distinct prime divisors of n. Let $b(n)$ be the largest integer value such that $2^{b(n)} | p - 1$, for every prime number p such that $p|n$.
(a) Let

$$C(n) = \left\{ a \in \mathbb{U}_n : a^{2^m \cdot l} \equiv \pm 1 \pmod{n} \right\} \tag{11.36}$$

Show that if $a \in MRnw(n)$ then $a \in C(n)$.
(b) Furthermore show that $MRnw(n) \subseteq C(n)$.

Solution.
Both (a) and (b) follow trivially from the prior discussion. ■

Exercise 1036.

Show the following.

Proposition 11.6 **Auxiliary result 1**

Let $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ be the prime decomposition of composite n, let $d(n) = r$ indicate the number of distinct prime divisors of n. Furthermore, let $n-1 = 2^k l$, where l is an odd integer. Let $y \in \mathbb{Z}$ such that $\gcd(y,n) = 1$ and let $b \in \mathbb{Z}_+$.

(a) Then, the following modular equation

$$x^v \equiv y \pmod{p_i^{a^i}}, \tag{11.37}$$

has a solution x mod $p_i^{a_i}$ if

$$d_i = \gcd(v, \phi(p_i^{a_i})) | ind_{g_i}(y),$$

and if there is at least one solution, conclude then that the number of solutions is d_i.

(b) Using CRT then the number of solutions of

$$x^v \equiv y \pmod{n}, \tag{11.38}$$

is given by the following expression

$$\prod_{i=1}^{r} d_i = \prod_{i=1}^{r} \gcd(v, \phi(p_i^{a_i})) = \prod_{i=1}^{r} \gcd(v, (p_i^{a_i-1}(p_i-1))). \tag{11.39}$$

We are going to use the following result.

Theorem 11.5 **Linear congruences with $d > 1$**

Let $d = \gcd(a,n) > 1$. If $d|b$ the linear congruence below has exactly d solutions mod n.

$$ax \equiv b \pmod{n}. \tag{11.40}$$

Solution.

(a) Let g_i be a generator of $\mathbb{U}_{p_i^{a_i}}$. Let $ind_{g_i}(x)$ and $ind_{g_i}(y)$ be the indices of x, y respectively.

$$\begin{aligned} x^v &\equiv y \pmod{n} \Leftrightarrow \\ x^v &\equiv y \pmod{p_i^{a_i}} \ \forall i = 1, \dots, r \Leftrightarrow \\ g_i^{ind_{g_i}(x)v} &\equiv g_i^{ind_{g_i}(y)} \pmod{p_i^{a_i}} \Leftrightarrow \\ ind_{g_i}(x)v &\equiv ind_{g_i}(y) \pmod{\phi(p_i^{a_i})} \end{aligned} \tag{11.41}$$

Viewing the last equation Eq.(11.41) as a modular equation and invoking the results associated with Eq.(11.5) we obtain the following. A solution for Eq.(11.41) exists if and only if

$$d_i | ind_{g_i}(y),$$

where

$$d_i = \gcd(v, \phi(p_i^{a_i})).$$

If one solution exists, then the number of solution mod $p_i^{a_i}$ is equal to d_i.

(b) Follows by the CRT and the fact that $\phi(p_i^{a_i}) = (p_i^{a_i-1}(p_i-1))$. ■

Exercise 1037.

Show the following.

Proposition 11.7 **Auxiliary result 2**

Let $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ be the prime decomposition of composite n, let $d(n) = r$ indicate the number of distinct prime divisors of n. Furthermore, let $n-1 = 2^k l$, where l is an odd integer. Let $b(n)$ be the largest integer value such that $2^{b(n)} | p-1$, for every prime number p such that $p|n$.
We reintroduce Eq. (11.36).

$$C(n) = \left\{ a \in \mathbb{U}_n : a^{2^m \cdot l} \equiv \pm 1 \pmod{n} \right\} \tag{11.42}$$

Show the following.

$$c = |C(n)| = 2 \cdot \prod_{i=1}^{r} 2^{(b(n)-1)} \gcd(p-1, l) = 2 \cdot 2^{(b(n)-1) \cdot d(n)} \cdot \prod_{i=1}^{r} \gcd(p-1, l). \tag{11.43}$$

Solution.
For $b(n)$ let us define $v = 2^{b(n)-1} \cdot l$. Equation (11.36) leads us to count the number of a of $C(n)$ that satisfy modular equation

$$a^v \equiv 1 \pmod{n} \Leftrightarrow a^v \equiv 1 \pmod{p_i^{a_i}} \quad \forall i = 1, 2, \dots, r \tag{11.44}$$

or modular equation

$$a^v \equiv -1 \pmod{n} \Leftrightarrow a^v \equiv -1 \pmod{p_i^{a_i}} \quad \forall i = 1, 2, \dots, r \tag{11.45}$$

Case 1: Number of solutions of Eq.(11.44). We start with the former. Eq.(11.44) by Eq.(11.37) has a number of solutions equal to

$$d_i = \gcd(v, \phi(p_i^{a_i})),$$

as long as

$$d_i | ind_{g_i}(y) = d_i | ind_{g_i}(1)$$

which is the case for $ind_{g_i}(1) = \phi(p_i^{a_i})$. Then

$$\begin{aligned} \forall i = 1, 2, \dots r \quad : \gcd(v, \phi(p_i^{a_i})) &= \gcd(v, p_i^{a_i - 1}(p_i - 1)) \\ &= \gcd(v, (p_i - 1)) \\ &= \gcd(2^{b(n)-1} \cdot l, (p_i - 1)) \\ &= 2^{b(n)-1} \cdot \gcd(l, (p_i - 1)). \end{aligned} \tag{11.46}$$

The term $p_i^{a_i - 1}$ was removed from the first equation above, because $p_i \nmid v$. If it were $p_i | v$ then since $v | n-1$ we would have had $p_i | n-1$. Since $p_i | n$ the last two divisibility results would lead to $p_i | 1$ i.e. $p_i = 1$. This violates the assumption of the unique factorization of n where p_i are prime numbers and in fact greater than two.
By the CRT and combining for $i = 1, 2, \dots, r$, Eq. (11.46) we obtain that for modular equation

$$a^v \equiv 1 \pmod{n}$$

the number of its solutions is as follows.

$$\prod_{i=1}^{r} 2^{b(n)-1} \cdot \gcd(l, (p_i - 1)) = 2^{(b(n)-1)r} \cdot \prod_{i=1}^{r} \gcd(l, (p_i - 1)) = 2^{(b(n)-1)d(n)} \cdot \prod_{i=1}^{r} \gcd(l, (p_i - 1)).$$

Case 2: Number of solutions of Eq.(11.45). We continue with the latter Eq.(11.45) to establish the number of solutions of

$$a^v \equiv -1 \pmod{n}$$

by establishing the number of solutions of

$$a^v \equiv -1 \pmod{p_i^{a_i}} \quad \forall i = 1,2,\ldots,r.$$

Note that the latter implies

$$a^{2v} \equiv 1 \pmod{p_i^{a_i}} \quad \forall i = 1,2,\ldots,r.$$

Therefore we find the number of solutions of

$$a^{2v} \equiv 1 \pmod{p_i^{a_i}} \quad \wedge \quad a^v \not\equiv 1 \pmod{p_i^{a_i}} \quad \forall i = 1,2,\ldots,r.$$

We start with the former. The d_i is slightly different.

$$d_i = \gcd(2v, \phi(p_i^{a_i})),$$

We note that we need

$$d_i | ind_{g_i}(y) = d_i | ind_{g_i}(1)$$

which is the case as before for Case 1. Then

$$\begin{aligned} \forall i = 1,2,\ldots r \quad : \gcd(2v, \phi(p_i^{a_i})) &= \gcd(2v, p_i^{a_i-1}(p_i-1)) \\ &= \gcd(2v, (p_i-1)) \\ &= \gcd(2^{b(n)} \cdot l, (p_i-1)) \\ &= 2^{b(n)} \cdot \gcd(l, (p_i-1)). \end{aligned} \tag{11.47}$$

$$a^{2v} \equiv 1 \pmod{p_i^{a_i}} \quad \wedge \quad a^v \not\equiv 1 \pmod{p_i^{a_i}} \quad \forall i = 1,2,\ldots,r.$$

Therefore the number of solutions of

$$a^{2v} \equiv 1 \pmod{p_i^{a_i}} \quad \wedge \quad a^v \not\equiv 1 \pmod{p_i^{a_i}} \quad \forall i = 1,2,\ldots,r$$

is equal to

$$2^{b(n)} \cdot \gcd(l, (p_i-1)) - 2^{b(n)-1} \cdot \gcd(l, (p_i-1)) = 2^{b(n)-1} \cdot \gcd(l, (p_i-1)),$$

just like case 1. This concludes the case and the result. ■

Exercise 1038.

Show the following.

Proposition 11.8 **Auxiliary result 3**

Let $n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r}$ be the prime decomposition of composite n, and let $d(n) = r$ indicate the number of distinct prime divisors of n. Furthermore, let $n-1 = 2^k l$, where l is an odd integer. Let $b(n)$ be the largest integer value such that $2^{b(n)} | p-1$, for every prime number p such that $p|n$.
For $C(n) = \left\{ a \in \mathbb{U}_n : a^{2^m \cdot l} \equiv \pm 1 \pmod{n} \right\}$ Then, we have the following.

$$c = |C(n)| \leq \frac{\phi(n)}{4}, \tag{11.48}$$

for all odd composite $n > 10$.

Solution.
From Eq.(11.43) we have the following.

$$c = |C(n)| = 2 \cdot \prod_{i=1}^{r} 2^{(b(n)-1)} \gcd(p-1, l) \tag{11.49}$$

$$\frac{\phi(n)}{|C(n)|} = \frac{1}{2} \cdot \prod_{i=1}^{r} \frac{p_i^{a^i - 1}(p_i - 1)}{2^{(b(n)-1)} \gcd(p-1, l)} \tag{11.50}$$

We note a few things related to Eq.(11.50). Since $2^{b(n)-1} | 2^{b(n)}$ and $2^{b(n)} | p-1$ for each i, we have that $\frac{(p_i-1)}{2^{b(n)-1}}$ is an integer and also an even number. Moreover $\gcd(l, p_i - 1) | p_i - 1$ and $\gcd(l, p_i - 1)$ is an odd number, since l is odd. Thus the product term of Eq.(11.50) is an integer and a multiple of two and multiplied by the outside $1/2$ still yields an integer. We now perform a case analysis.
Case 1: $r = d(n) \geq 3$**.** Then

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot 2 \cdot 2 \cdot 2 = 4.$$

Case 2: $r = d(n) = 2$**, and** n **is not squarefree.** Then for some p_i we have $a_i \geq 2$. In the numerator of Eq.(11.50) $p_i^{a_i - 1}$ contributes a $p_i^{a_i - 1} \geq p_i \geq 3$.

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot (3 \cdot 2) \cdot 2 = 6.$$

Case 3: $r = d(n) = 2$**, but** n **is squarefree.** Let $n = pq$ where $p < q < n$. We have that $2^{b(n)} | p-1$ and $2^{b(n)} | q-1$. We distinguish two subcases.
Case 3a: $2^{b(n)+1} | q-1$**.** This means $q - 1 = 4 \cdot 2^{b(n)-1} \cdot R$. Then we have as follows.

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot 2 \cdot \frac{q-1}{2^{b(n)-1} \gcd(l, q-1)} \geq \frac{1}{2} \cdot 2 \cdot 4 = 4.$$

Case 3b: $2^{b(n)+1} \nmid q-1$ **but** $2^{b(n)1} | q-1$**.** Then we have as follows.

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot 2 \cdot \frac{q-1}{2^{b(n)-1} \gcd(l, q-1)} \frac{1}{2} \cdot 2 \cdot \frac{2 \cdot Q}{\gcd(l, q-1)} \tag{11.51}$$

Note that $n - 1 = 2^k l$, where $b(n) \leq k$. Moreover, $q - 1 = 2^{b(n)} Q$, where Q can be even or odd. Furthermore,

$$n - 1 = pq - 1 = p(q-1) + (p-1) \Rightarrow n - 1 \equiv p - 1 \pmod{q-1} \Rightarrow q - 1 \nmid n - 1.$$

The latter implies that there is a prime divisor of $q-1$ that does not divide $n-1$. Since $b(n) \leq k$, this divisor is not a two. So it must be an odd number and it is at least 3. Therefore $q-1 = 2Q = 2 \cdot 3 \cdot P$. Therefore we refine the previous bound of Eq.(11.51) as follows.

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot 2 \cdot \frac{2 \cdot Q}{\gcd(l, q-1)} \geq \frac{1}{2} \cdot 2 \cdot \frac{2 \cdot 3 \cdot P}{\gcd(l, q-1)} \geq 6. \tag{11.52}$$

Case 4: $d(n) = 1$ **and** n **is not squarefree i.e.** $n = p^a$**, for some** $a \geq 2$ **and** $p \geq 3$**.** A minimal such value is for $a = 2$ and $p = 3$ and it is $n = 3^2 = 9$. We then have he following.

$$\frac{\phi(n)}{|C(n)|} \geq \frac{1}{2} \cdot \frac{p^{a-1}(p-1)}{2^{b(n)-1} \gcd(l, p-1)}$$

Note that $p - 1 = 2^k \cdot l$. Moreover $b(n) = k$. Then $\gcd(l, p-1) = \gcd(l, 2^k \cdot l) = l$. Therefore

$$\frac{1}{2} \cdot \frac{p^{a-1}(p-1)}{2^{b(n)-1} \gcd(l, p-1)} = \frac{1}{2} \cdot \frac{p^{a-1} \cdot 2^k \cdot l}{2^{k-1} l} = p^{a-1}.$$

If $p = 3$ and $a = 2$ this p^{a-1} is a 3. We discard this possibility by requiring $n > 9$ and for composite and odd n this becomes $n > 10$. In all other cases p^{a-1} is at least 5. ■

11.5 Prime number theorem

Exercise 1039.

Let $\pi(n)$ be the number of prime numbers less than or equal to n. Then $\pi(n) \Rightarrow \infty$ as $n \Rightarrow \infty$. It is $\pi(n) \approx n/\ln n$, or equivalently the following holds.

$$\lim_{n\to\infty} \frac{\pi(n)}{n/\ln n} = 1. \tag{11.53}$$

Show the result known as the prime number theorem.

Solution.
Missing. ■

Exercise 1040.

Let $\pi(n)$ be the number of prime numbers less than or equal to n. Let p_n be the n-th prime number. Show the following

$$\lim_{n\to\infty}\frac{\pi(n)}{n/\ln n}=1\Leftrightarrow\lim_{n\to\infty}\frac{p_n}{n\ln n}=1 \tag{11.54}$$

Solution.
$\Rightarrow$. We start with Eq.(11.53).

$$\begin{aligned}
\lim_{n\to\infty}\frac{\pi(n)}{n/\ln n}=\lim_{n\to\infty}\frac{\pi(n)\ln n}{n} &= 1\\
\lim_{n\to\infty}\ln\left(\frac{\pi(n)\ln n}{n}\right) &= \ln 1\\
\lim_{n\to\infty}\left(\ln(\pi(n))+\ln(\ln n)-\ln n\right) &= 0\\
\lim_{n\to\infty}\ln n\cdot\left(\frac{\ln\pi(n)}{\ln n}+\frac{\ln(\ln n)}{\ln n}-1\right) &= 0\\
\lim_{n\to\infty}\ln n\cdot\left(\frac{\ln\pi(n)}{\ln n}-1\right) &= 0\\
\lim_{n\to\infty}\ln\pi(n) &= \lim_{n\to\infty}\ln n\\
\lim_{n\to\infty}\frac{\ln\pi(n)}{\ln n} &= 1
\end{aligned} \tag{11.55}$$

We multiply Eq.(11.53) with Eq.(11.55) to derive the following.

$$\begin{aligned}
\lim_{n\to\infty}\frac{\ln\pi(n)}{\ln n}=1 \quad\wedge\quad \lim_{n\to\infty}\frac{\pi(n)}{n/\ln n}=1\Leftrightarrow\\
\lim_{n\to\infty}\frac{\pi(n)\cdot\ln\pi(n)}{n} &= 1
\end{aligned} \tag{11.56}$$

In the latter Eq.(11.56) we substitute $n=p_n$ and $\pi(n)=\pi(p_n)=n$. We then obtain the following.

$$\begin{aligned}
\lim_{n\to\infty}\frac{\pi(n)\cdot\ln\pi(n)}{n} &= 1\Leftrightarrow\\
\lim_{n\to\infty}\frac{n\cdot\ln n}{p_n} &= 1.
\end{aligned} \tag{11.57}$$

This concludes this case, after an inversion.
$\Leftarrow$. Shown similarly. ■

11.5.1 Solovay-Strassen and Miller-Rabin failure probability bounds

Exercise 1041.

Both in the t-round probabilistic version of the Solovay-Strassen test and probabilistic version of the Rabin-Miller test we derive bounds on the probability of

$$P_{SolStr}(E_1/E_2) \leq \frac{1}{2^t},$$

$$P_{RabMil}(E_1/E_2) \leq \frac{1}{4^t},$$

where

$$E_1 : \text{event that } t \text{ runs of the algorithm produce NO witnesses,}$$

$$E_2 : n \text{ is a composite number,}$$

and of course

$$E_2' : n \text{ is a prime number.}$$

However we are interested in the probability

$$P_{SolStr}(E_2/E_1),$$

and

$$P_{RabMil}(E_2/E_1),$$

instead.

Use Bayes theorem to conclude ([11]) the following.

$$P_{SolStr}(E_2'/E_1) \geq 1 - \frac{\ln n - 1}{2^t}$$

and

$$P_{RabMil}(E_2'/E_1) \geq 1 - \frac{\ln n - 1}{4^t},$$

or equivalently

$$P_{SolStr}(E_2/E_1) \leq \frac{\ln n - 1}{2^t}$$

and

$$P_{RabMil}(E_2/E_1) \leq \frac{\ln n - 1}{4^t},$$

Solution.

Bayes theorem simple form. The simple form of Bayes theorem for two event A and B with $P(B) \neq 0$ takes the following form.

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}.$$

A proof is based on conditional probability. Therefore for events A, B we have the following.

$$P(A \cap B) = P(A/B)P(B), \qquad P(A \cap B) = P(B/A)P(A).$$

Equating the right-hand sides and solving for $P(A/B)$ given $P(B) \neq 0$ gives the desired result.

Bayes theorem for two mutually exclusive and exhaustive events A and A'. We denote A' the complement of A with $A \cap A' = \emptyset$ and $A \cup A' = \Omega$ that is A and A' are mutually exclusive and exhaustive. Then

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/A')P(A')}. \tag{11.58}$$

Likewise we can also have the following.

$$P(A'/B) = \frac{P(B/A')P(A')}{P(B/A)P(A)+P(B/A')P(A')}. \tag{11.59}$$

We map A, A', B to the events E_2, E_2', E_1.

$$\begin{aligned} A &\mapsto E_2, \quad n \text{ is Composite event,} \\ A' &\mapsto E_2', \quad n \text{ is Prime event,} \\ B &\mapsto E_1. \quad \text{NoWitness event in } t \text{ rounds.} \end{aligned}$$

Lemma 11.2

For complementary events $E_1 \cap E_2$ and $E_1 \cap E_2'$ we have the following.

$$P(E_2/E_1) = 1 - P(E_2'/E_1) \tag{11.60}$$

For complementary events $E_1 \cap E_2$ and $E_1 \cap E_2'$ we have

$$\begin{aligned} P(E_1) &= P(E_1 \cap E_2) + P(E_1 \cap E_2') \\ \frac{P(E_1)}{P(E_1)} &= \frac{P(E_1 \cap E_2) + P(E_1 \cap E_2')}{P(E_1)} \\ 1 &= \frac{P(E_1 \cap E_2)}{P(E_1)} + \frac{P(E_1 \cap E_2')}{P(E_1)} \\ 1 &= P(E_2/E_1) + P(E_2'/E_1), \end{aligned}$$

and after solving for $P(E_2/E_1)$ the lemma follows.
Furthermore we have the following for the Solovay-Strassen algorithm

$$P(E_1/E_2) \leq \frac{1}{2^t}. \tag{11.61}$$

By Equation 11.59 we have the following with the substitutions obtained through the three mappings for A, A' and B.

$$P(E_2'/E_1) = \frac{P(E_1/E_2')P(E_2')}{P(E_1/E_2)P(E_2)+P(E_1/E_2')P(E_2')}. \tag{11.62}$$

By the prime number theorem, quoting the density of primes, we have the following.

$$P(E_2') \approx \frac{1}{\ln n}. \tag{11.63}$$

Furthermore,

$$P(E_2) = 1 - P(E_2') \approx 1 - \frac{1}{\ln n}. \tag{11.64}$$

Moreover,

$$P(E_1/E_2') = \frac{P(E_1 \cap E_2'}{P(E_2')} = 1, \tag{11.65}$$

since no witness for the compositeness of n is to be generated for n a prime number. Therefore Eq. (11.62) by way of Eq.(11.63), Eq.(11.64), Eq.(11.65 yields the following.

$$\begin{aligned} P(E_2'/E_1) &= \frac{P(E_1/E_2')P(E_2')}{P(E_1/E_2)P(E_2)+P(E_1/E_2')P(E_2')} \\ &= \frac{1\cdot\frac{1}{\ln n}}{P(E_1/E_2)\cdot(1-\frac{1}{\ln n})+1\cdot\frac{1}{\ln n}} \\ &= \frac{1}{P(E_1/E_2)\cdot(\ln n-1)+1} \text{by way of Eq. (11.61)} \\ &\geq \frac{1}{1+\frac{\ln n-1}{2^t}} \\ &\geq 1-\frac{\ln n-1}{2^t}, \end{aligned} \tag{11.66}$$

where in the last equation one can use $1/(1+x)\geq 1-x$ for $0<x<1$. Therefore one can also write

$$P(E_2'/E_1)\geq 1-\frac{\ln n-1}{2^t}\geq 1-\frac{\ln n}{2^t}, \tag{11.67}$$

and by using Lemma 11.2 one can obtain the following.

$$P(E_2/E_1)\leq \frac{\ln n-1}{2^t}\leq \frac{\ln n}{2^t}. \tag{11.68}$$

The correspond results for the Random-Miller probabilistic primality test then become as follows.

$$P(E_2'/E_1)\geq 1-\frac{\ln n-1}{4^t}\geq 1-\frac{\ln n}{4^t}, \tag{11.69}$$

$$P(E_2/E_1)\leq \frac{\ln n-1}{4^t}\leq \frac{\ln n}{4^t}. \tag{11.70}$$

■

Bibliography

[1] M. Abramowitz and I. A. Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. New York: Dover Publications. Ninth printing.

[2] D. Angluin and L. G. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. Journal of Computer and System Sciences. Volume 18, Issue 2, 1979, Pages 155-193.

[3] E. Bach. Explicit bounds for primality testing and related problems. Mathematics of Computation, Vol 55, No. 191 (Jul., 1999), pp. 355-380.

[4] B. Béla Bollobás. Random Graphs. Academic Press 1985.

[5] C. Berge. Two theorems in graph theory. Proc. Nat. Ac. Sciences., U.S.A., 43, 1957, 842.

[6] C. Berge. Sur le couplage maximum d'un graphe. C. R. Acad. Sciences, 247, 1958, 258-259.

[7] C. Berge. "Alternating chains", in: Graph Theory and Computing (R. C. Read), Academic Press, 1972.

[8] C. Berge. "Graph and hypergraphs", North Holland, 1973.

[9] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics. 23 (4):493-507, 1952.

[10] V.Chvátal. The tail of the hypergeometric distribution. Discrete Mathematics. 25(3):285-287, 1979, Elsevier.

[11] K. Conrad. The Solovay-Strassen test.
https://kconrad.math.uconn.edu/blurbs/ugradnumthy/solovaystrassen.pdf [Accessed: 2026/05/04]

[12] K. Conrad. The Miller-Rabin test.
https://kconrad.math.uconn.edu/blurbs/ugradnumthy/millerrabin.pdf [Accessed: 2026/05/05]

[13] H. Cramér. Sus un nouveau thèoréme-limite de la théorie des probabilités. Actualités Scientifiques et Industrielles. No 736. Paris 1938.

[14] R. Crandall and C. Pomerance. Prime Numbers: A computational perspective. Second edition. Springer, 2005.

[15] E. Egerváry. Über kombinatorische eigenschaften von matrizen, Mat. Fiz. Lapok 38, 16-28, 1931.

[16] P. Erdos and J. Spencer. Probabilistic methods in combinatorics. Academic Press. New York, 1974.

[17] P. Erdos and T. Gallai. On maximal path and circuits of graphs. Acta Math. Acad. Sc. Hungar., 10, 1959, 337-356.

[18] W. Feller. An introduction to probability theory and its applications. Vol. 1. 3rd Edition. John Wiley & Sons. New York, 1968.

[19] W. Feller. An introduction to probability theory and its applications. Vol. 2. John Wiley & Sons. New York, 1971.

[20] G. Frobenius. Über zerlegbare determinanten, sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, 456-477, 1917.

[21] A. V. Gerbessiotis and L. G. Valiant. Direct bulk-synchronous parallel algorithms. Journal of Parallel and Distributed Computing. 22(1994), pp. 251-267, 1994.

[22] P. Hall. On representatives of subsets. J. London Math. Soc., 10, 1935, 26-30.

[23] W. Hoeffding. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association. 58 (301):13-30, March 1963.

[24] D. Kőnig. Gráfok és mátrixok. (Also as Graphen und matrizen.) Matematikai és Fizikai Lapok, 38, 116-119, 1931.

[25] S. Kullback and R. A. Leibler. On information and sufficiency. The Annals of Mathematical Statistics. 22 (1):79-86, 1952.

[26] C. McDiarmid . On the method of bounded differences. In: Siemons J, ed. Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference. London Mathematical Society Lecture Note Series. Cambridge University Press; 1989:148-188.

[27] G. L. Miller. Riemann's hypthesis and tests for primality. Journal of computer and system sciences, 13, 300-317 (1976), Academic Press.

[28] L. Monier. Evaluation and comparison of two efficient probabilistic primality testing algorithms. Theoretical Computer Science, 12(1980), 97-108, North-Holland Publishing Company.

[29] National Institute of Standards and Technology. https://dlmf.nist.gov/4.5

[30] National Institute of Standards and Technology. https://dlmf.nist.gov/4.6

[31] R. Z. Norman and M. O. Rabin. An algorithm for a minimum cover of a graph. Proc. Amer. Math. Soc., 10, 1959, 315-319.

[32] M. O. Rabin. Probabilistic algorithm for testing primality. Journal of Number Theory, 12, 128-138 (1980), Academic Press.

[33] R. Solovay and V. Strassen. A fast Monte-Carlo test for primality. SIAM Journal of computing, 6 (1977), 84-85.

[34] Wikipedia. https://en.wikipedia.org/wiki/Floor_and_ceiling_functions [Accessed: 2026/05/28]