Probability for computer science: inequalities and their bounds

Version 2.0 (June 23, 2026)

Alexandros V. Gerbessiotis

CS Department
NJIT
Newark, NJ 07102.
Email: alexg@njit.edu

Printed on June 23, 2026

Preface

This monograph provides a basic review on probability topics for computing students. Students in other disciplines such as engineering, applied mathematics and statistics, and data science will also find the topics interesting.

The discussion on probability starts with experiments and random processes, sample spaces, event spaces and probability spaces to a level relevant and applicable to computer science.

It then reviews random variables, distributions, and their properties. With emphasis on discrete probability some standard distributions and their properties are examined in the form of examples.

After an introduction to the concept of a moment generating function, probability inequalities are discussed.

A comprehensive survey on Chernoff's bounds and its derivations and Hoeffding's bounds is also included. Proofs are included to make this monograph self contained.

Finally a collection of problems with some sample solutions are included. Again, the emphasis is on computer science relevant topics, but problems on random graphs and Ramsey numbers are also included into the mix.

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.

Contents

I Introduction **1**

1 Experiments and trials **3**
1.1 Experiments and trials . . . 3
1.2 Events . . . 3
1.3 Event spaces and measurable spaces . . . 4
1.3.1 σ-algebras . . . 4
1.3.2 Measurable spaces . . . 4
1.3.3 Topologies . . . 5
1.3.4 Measures . . . 5
1.4 Probability space . . . 6
1.4.1 Probability measure . . . 6
1.4.2 Probability space defined . . . 6
1.4.3 Finite probability space . . . 7

2 Probability of events **9**
2.1 Probability of events . . . 9
2.2 Conditional probabilities . . . 12

II Random variables and distributions **15**

3 Random variables **17**
3.1 Random variables . . . 17
3.2 Discrete random variables . . . 18
3.2.1 Probability mass function (p.m.f.) . . . 19
3.2.2 Composition of random variables . . . 21
3.2.3 Expectation . . . 21
3.2.4 Joint probability . . . 23
3.2.5 Independent random variables . . . 25
3.2.6 Variance . . . 26
3.2.7 Cumulative distribution function . . . 28
3.2.8 Probability generating function . . . 29
3.2.9 Miscellanea . . . 29
3.3 Continuous random variables . . . 31
3.3.1 Cumulative distribution function . . . 31
3.3.2 Probability density function (p.d.f.) . . . 32
3.3.3 Expectation of a continuous r.v. . . . 32
3.3.4 Independent random variables: continuous case . . . 32
3.3.5 Variance of a continuous r.v. . . . 33

3.3.6 Median and other quartiles . . . 35
3.3.7 Logarithmic and other inequalities . . . 35
3.3.8 Composition of random variables . . . 38
3.4 Conditional expectation . . . 38
3.5 More on expectations . . . 40

4 Standard distributions and random variables 43
4.1 Standard distributions . . . 43
4.1.1 Uniform distribution . . . 43
4.1.2 Bernoulli distribution . . . 43
4.1.3 Geometric distribution . . . 44
4.1.4 Binomial distribution . . . 45
4.1.5 Normal distribution . . . 47
4.2 Exercises . . . 48

5 Generating functions 49
5.1 Generating functions . . . 49
5.1.1 Probability generating function . . . 49
5.1.2 Moments . . . 50
5.1.3 Moment generating function . . . 50
5.1.4 Some examples . . . 51
5.2 Indicator random variable . . . 52
5.3 Characteristic function . . . 53

III Inequalities in probability theory 55

6 Probability inequalities 57
6.1 Probability inequalities . . . 57
6.1.1 Markov's inequality . . . 58
6.1.2 Cauchy-Schwartz inequality . . . 59
6.1.3 Tsebyshev's or Tchebychef's inequality . . . 61
6.1.4 Jensen's inequality . . . 63
6.1.5 Law of large numbers . . . 64
6.1.6 Holder's inequality . . . 65
6.1.7 Liapunov's inequality . . . 65
6.1.8 Minkowski's inequality . . . 65
6.1.9 Tails of the binomial distribution . . . 66

7 Hoeffding's method 69
7.1 Hoeffding's inequality . . . 69
7.2 Derived right tails . . . 69
7.3 Derived left tails . . . 75
7.4 Derived concentration bounds . . . 76

8 Chernoff's method 77
8.1 Chernoff's inequality . . . 77
8.2 Derived right tails . . . 79
8.3 Derived left tails . . . 84
8.4 Derived concentration bounds . . . 87

9 Combinatorial inequalities **89**
9.1 Combinatorial inequalities . . . 89
9.2 Series . . . 93
9.3 Exponential inequalities . . . 94
9.4 Logarithmic and other inequalities . . . 95
9.5 Combinatorial equalities . . . 99

IV Probability problems **103**

10 Introductory problems **105**
10.1 Entropy . . . 105
10.2 Spaces . . . 108

11 Probability of events **111**
11.1 Events and their properties . . . 111
11.2 Toys: dies, coins and cards . . . 117
11.2.1 Throw or roll a die or dice . . . 118
11.2.2 Toss a coin . . . 122
11.2.3 Deck of cards . . . 133
11.3 Conditional probabilities . . . 136
11.4 Independence of events . . . 142

12 Random variables **159**
12.1 Discrete random variables . . . 159
12.2 Continuous random variables . . . 166
12.3 Birthdays . . . 178
12.4 Miscellanea . . . 183
12.5 Permutations and Derangements . . . 195
12.6 Bin and Balls . . . 202
12.6.1 ... or not . . . 206
12.7 Collecting coupons . . . 217

13 Distributions **221**
13.1 Distributions . . . 221
13.1.1 Binomial process and Bernoulli trials . . . 221
13.2 Randomness . . . 231
13.3 Ramsey numbers . . . 242
13.4 Van der Waerden . . . 247

14 Generating functions **249**
14.1 Generating functions . . . 249

15 Probability inequalities **267**
15.1 Probability inequalities . . . 267

16 Miscellaneous problems **277**
16.1 Hashing . . . 277
16.2 Hashing . . . 287
16.3 Miscellanea . . . 291
16.4 Random graphs . . . 296
16.5 Random matrices . . . 308

16.6 Puzzle 312

Part I

Introduction

Chapter 1

Experiments and trials

1.1 Experiments and trials

A trial or elementary experiment has outcomes. An experiment that is not elementary can be considered a sequence of individual trials (or elementary experiments). 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 involve the rolling of a die say three times. 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 1.1 **Experiment or trial**

An experiment or trial is any procedure that can be repeated and generate 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 1.2 **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 1.3 **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.

1.2 Events

Definition 1.4 **Event**

An event is a set of outcomes that is a subset of sample space S.

Sometimes we refer to 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 trial. Event E_1, where $E_1 = \{HHT, HTH, THH\}$ 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 1.5 **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.

In the remainder, a countable set is meant to be a finite set or a countably infinite set.

1.3 Event spaces and measurable spaces

Definition 1.6 **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.

1.3.1 σ-algebras

Definition 1.7 **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 $A_1, A_2, \ldots, A_i$, $A_i \in T$ 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).

1.3.2 Measurable spaces

Definition 1.8 **(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.

1.3.3 Topologies

Definition 1.9 **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 1.3.1

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 1.3.2

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).

1.3.4 Measures

Definition 1.10 **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, $m(\cup_i A_i) = \sum_i m(A_i)$.

1.4 Probability space

A probability measure denoted as P is a measure where $\mathbb{R}$ 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.

1.4.1 Probability measure

Definition 1.11 **Probability measure P**

Consider a measurable space (S,T) of a sample space of outcomes and a set of events T that is a σ-algebra. 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 occur 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 $\{1,2,3,4,5,6\}$.

1.4.2 Probability space defined

Definition 1.12 **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 countably additive probability measure on T with $P(S) = 1$.

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 outcomes (elementary events) of A. 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.

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 1.1

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 $\lim_{i \to \infty} P(A_i) = P(A)$.*

1.4.3 Finite probability space

Definition 1.13 **Finite probability space**

Let S be a finite sample space $S = \{s_1, \ldots, s_n\}$. A probability model or finite probability space is obtained by assigning to each point $s_i \in S$ a real number p_i (or $p(i)$), the probability of s_i, that satisfies the following properties.

- Each p_i is nonegative $p_i \geq 0$, and
- the sum of p_i is one i.e. $\sum_{1 \leq i \leq n} p_i = p_1 + p_2 + \ldots + p_n = 1$.

Definition 1.14 **Event probability**

The probability of an event A denoted as $P(A)$ is the sum of the probabilities of the elements of A.

Definition 1.15 **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 $\sum_{s \in S} p(s) = 1$.

Proposition 1.1

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 $\sum_i P(\{s_i\}) = 1$.

Sometimes we may refer to it as finite probability space but finite refers to a finite S whereas discrete refers to an S that is countable (finite or countably infinite).

Note also the difference between $P(.)$ and $p(.)$. The latter is defined on outcomes (or elementary events). The former is defines of events that are subsets of sample space S. An elementary event is also an event and therefore, $P(\{s\}) = p(s)$.

Definition 1.16 **Event probability properties**

The probability function P defined on the class of events of a finite probability space has the following properties.

- For every event A, we have $1 \geq P(A) \geq P(\emptyset) = 0$,
- if events A, B are mutually exclusive then $P(A \cup B) = P(A) + P(B)$, and
- $P(S) = 1$.

Definition 1.17 **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 1.18 **High probability event**

We say that 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.

Chapter 2

Probability of events

2.1 Probability of events

Corollary 2.1

Let A be an event and the probability function P defined on the class of events of a finite probability space. Then $P(A^{\complement}) = 1 - P(A)$.

Lemma 2.1

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

$$B_i = A_i - (A_1 \cup \ldots \cup A_{i-1})$$

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 2.2 **Independent Events**

Two events A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

Theorem 2.1 **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

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

The latter equality is also known as the inclusion-exclusion property or principle. It can be generalized to three or more events (with a proof by induction). It is also straightforwars to derive the following.

Corollary 2.2

Given two events A, B we have

$$P(A \cup B) \leq P(A) + P(B)$$

Theorem 2.2 **Inclusion-Exclusion of three events**

Given three events A_1, A_2, A_3 we have the following

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_1 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Definition 2.1

Let S be a sample space and $A_i \subseteq S$ events. For each non empty subjet $I \subseteq \{1, \ldots, n\}$ we define

$$A_I = \cap_{i \in I} A_i.$$

By default $A_0 = S$.

Theorem 2.3 **Inclusion-Exclusion**

We have

$$|A_1 \cup \ldots \cup A_n| = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |A_I|. = \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |\cap_{i \in I} A_i|.$$

Proof. By induction on n. For $n = 1$, trivially $|A_1| = |A_1|$. From n to $n+1$ we use the $n = 2$ case.

$$\begin{aligned} |\cup_{i=1}^{n+1} A_i| &= |(\cup_{i=1}^{n} A_i) \cup A_{n+1}| \\ &= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |(\cup_{i=1}^{n} A_i) \cap A_{n+1}| \\ &= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |\cup_{i=1}^{n} (A_i \cap A_{n+1})| \\ &= \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |\cap_{i \in I} A_i| + |A_{n+1}| - \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |(\cap_{i \in I} A_i \cap A_{n+1})| \\ &= \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |\cap_{i \in I} A_i| + |A_{n+1}| - \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} |(\cap_{i \in I \cup \{n+1\}} A_i)| \\ &= \sum_{I \subseteq \{1,\ldots,n+1\}, n+1 \notin I} (-1)^{|I|+1} |\cap_{i \in I} A_i| + \sum_{I \subseteq \{1,\ldots,n+1\}, n+1 \in I} (-1)^{|I|+1} |(\cap_{i \in I} A_i)| \\ &= \sum_{I \subseteq \{1,\ldots,n+1\}} (-1)^{|I|+1} |\cap_{i \in I} A_i| \end{aligned}$$

■

By DeMorgan's Law the $\cap$ and $\cup$ can be swapped without affecting the result otherwise. When needed we would like to calculate

$$|A| - \sum_{I \subseteq \{1,\ldots,n+1\}} (-1)^{|I|+1} |\cap_{i \in I} A_i|,$$

or

$$|A| - \sum_{I \subseteq \{1,\ldots,n+1\}} (-1)^{|I|+1} |\cup_{i \in I} A_i|.$$

Example 2.1.1

For a sample space S, and $s \in S$, we call $\{s\}$ an elementary event. Moreover $\emptyset$ and S are also events. The impossible event or null event is $\emptyset$. If A, B are events so are $A \cap B$ and $A \cup B$ or $A^{\complement}$.

Example 2.1.2

An experiment is performed by throwing (tossing) a coin. The experiment is repeated twice. The combination of the results of two experiments is the experiment in question. The sample space S is $S = \{HH, HT, TH, TT\}$ 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, B is the outcome of the second experiment, where A, B is H or T. Event X is $X = \{HH, TT\}$ i.e. even number of H. Event Y is $Y = \{HH, HT, TH\}$ i.e. at least one H.

Example 2.1.3

An experiment is performed by tossing a coin. The experiment is repeated until an H is encountered. The sample space is infinite. Why? Because $S = \{T, TH, TTH, TTTH, TTTTH, \ldots\}$.

Example 2.1.4

(The singular form of dice is die.) An experiment is performed by throwing a (pair of) dice and records the number indicated at the top of the dice (opposite to the base that sits on a surface). Each die has six faces with six possible numbers, one on each face of a die. Then sample space has 36 outcomes

$$S = \{(a,b) : 1 \leq a, b \leq 6\}$$

Example 2.1.5

Deck of cards. A deck of card consists of 52 cards. There are 4 suits known as clubs(C), diamonds(D), hearts(H), and spades(S). Each suit contains 13 cards numbered 2 through 10, three face cards, jack (J), queen (Q), and king (K), and ace (A). The hearts and diamonds are red and spades and clubs are black.

Example 2.1.6

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 2.1.7

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 $P(A)$, $P(B)$, $P(A \cap B)$.

Proof. $P(A) = 13/52 = 1/4$. $P(B) = (3*4)/52 = 3/13$. $P(A \cap B) = 3/52$. ■

Example 2.1.8 **Uniform distribution**

$P(A) = c(A)/c(S)$ for all $A \subseteq S$.

2.2 Conditional probabilities

Theorem 2.4 **Conditional probability**

Let A, B be two events of finite probability space (S, Σ, 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 2.5 **Generalization of conditional probability**

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot P(A_4|A_1 \cap A_2 \cap A_3) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_{n-1})$$

Theorem 2.6

If sample space is partitioned into events $A_1, \ldots A_n$ then for some event A we have,

$$P(A) = \sum_{i=1}^{n} P(A/A_i)P(A_i).$$

We refer to the following as Bayes' rule.

Theorem 2.7

If sample space is partitioned into events $A_1, \ldots A_n$ then for some event A we have,

$$P(A_i/A) = \frac{P(A \cap A_i)}{P(A)} = \frac{P(A/A_i)P(A_i)}{\sum_i P(A/A_i)P(A_i)}$$

Moreover

Theorem 2.8

$P(A_1/A_2) = \frac{P(A_2/A_1)P(A_1)}{P(A_2)}$.

Proof. Note that $P(A_2/A_1)P(A_1) = P(A_1 \cap A_2)$. ■

Definition 2.2 **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 2.3

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 2.2.1

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 $P(A|B) = P(A \cap B)/P(B) = (2/36)/(5/36) = 2/5$.

Solution. 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 $P(A|B) = P(A \cap B)/P(B) = (2/36)/(5/36) = 2/5$. ■

When we talk about probability space (S,P), we remind ourselves that the absence of any reference to T as in probability space (S,T,P) implies that $T = \mathbb{P}(S)$,

Definition 2.3 **Independent repeated experiments**

Let (S,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.

$$P((s_1,\ldots,s_n)) = P(s_1)\ldots P(s_n).$$

Part II

Random variables and distributions

Chapter 3

Random variables

3.1 Random variables

The sample space S of an experiment or trial has outcomes that might be numbers 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 $T = 2^S = \mathbb{P}(S)$, 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 3.1 **Random variables: 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 3.2 **Random variables**

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 3.3 **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,

$$P(X \in A) = P(\{s \in S : X(s) \in A\}).$$

We have discrete and continuous random variables if the range of $R(S,X)$ is restricted on $\mathbb{Z}$ instead of $\mathbb{R}$.

3.2 Discrete random variables

The discussion below applies to a discrete random variable.

Definition 3.4 **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).

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 $E \subseteq T$.

Definition 3.5 **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 3.6 **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 3.2.1

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 $X : S \mapsto Z$, 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 3.2.2

For tossing a pair of dice sample space S has 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

$$X : S \to \mathbb{R} \quad \textbf{where } X(s) = X((a,b)) = a+b, \quad s \in S.$$

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. The range of X is $R(S,X)=\{1,2,3,4,5,6,7,8,9,10,11,12\}$. 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 $R(S,Y)=\{1,2,3,4,5,6\}$.

Example 3.2.3

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.

$$P(X\geq 1)=P(\{s\in S: X(s)\geq 1\})=P(X^{-1}(\{1,2\}))=P(\{(H,T),(T,H),(H,H)\})=3/4.$$

3.2.1 Probability mass function (p.m.f.)

Definition 3.7 **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 3.2.4

For Example 3.2.3

$$f_X : \{0,1,2\} \mapsto [0,1],$$

where

$$f_X(z) = P(X = z) = \binom{2}{z}\frac{1}{4}.$$

Example 3.2.5

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 3.1

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

$$\sum_{z \in Z} P(X = z) = P(\cup_{z \in Z}\{X = z\}) = P(S) = 1.$$

■

3.2.2 Composition of random variables

Definition 3.8 **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.

$$\forall s \in S : f(X, Y, \ldots, Z)(s) = f(X(s), Y(s), \ldots, Z(s)).$$

3.2.3 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 as $E(X)$ 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 3.9 **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 3.2

For a random variable X on Z and $f_X(z) = P(X = z)$, the following is given.

$$E(X) = E[X] = \sum_{s \in S} X(s)P(s) = \sum_{z \in Z} z f_X(z).$$

If the series (sum) converges, then X has finite expectation. If $\sum_z |z| f(z)$ diverges, then X has no finite expectation.

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 3.10

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 3.3

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

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} P(X > i) \\ &= P(X > 0) + P(X > 1) + P(X > 2) + P(X > 3) + \ldots \\ &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + \ldots \\ &+ P(X = 2) + P(X = 3) + P(X = 4) + \ldots \\ &+ P(X = 3) + P(X = 4) + \ldots \\ &+ P(X = 4) + \ldots \\ &= 1 \cdot P(X = 1) + 2 \cdot P(X = 3) + 3 \cdot P(X = 3) + 4 \cdot P(X = 4) \ldots \\ &= \sum_{i=1}^{\infty} iP(X = i). \end{aligned}$$

■

Example 3.2.6

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 3.11 **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 3.4

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.

$$\begin{aligned}
E(g(X)) &= \sum_{s\in S} g(X(s))P(s) \\
&= \sum_{s\in S} g(X(s))P(\{s\}) \\
&= \sum_{x} \sum_{s\in S, X(s)=x} g(X(s))P(\{s\}) \\
&= \sum_{x} \sum_{s\in S, X(s)=x} g(x)P(\{s\}) \\
&= \sum_{x} g(x) \sum_{s\in S, X(s)=x} P(\{s\}) \\
&= \sum_{x} g(x)P(X = x) \\
&= \sum_{x} g(x)f_X(x).
\end{aligned}$$

■

3.2.4 Joint probability

Definition 3.12

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 3.2.7

(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 3.5

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.

$$\begin{aligned}
E(W) &= \sum_w w f_W(w) \\
&= \sum_w w P(W=w) \\
&= \sum_w w P(g(X,Y)=w) \\
&= \sum_w w \sum_{x,y,w:w=g(x,y)} P(X=x,Y=y) \\
&= \sum_{x,y} \sum_{w:w=g(x,y)} w P(X=x,Y=y) \\
&= \sum_{x,y} g(x,y) P(X=x,Y=y).
\end{aligned}$$

■

3.2.5 Independent random variables

Definition 3.13 **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.

By induction, this can generalize to n random variables. Note that for two events A,B we defined the conditional probability of A on B as $P(A|B) = P(A\cap B)/P(B)$ and defined A is independent of B when $P(A) = P(A/B)$. Substituting the left-hand side of the latter for the left-hand side of the former we conclude $P(A/B) = P(A) = P(A\cap B)/P(B)$ from which we further derive $P(A\cap B) = P(A)P(B)$.

Theorem 3.6

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 3.2.8

(Dice tossing continued.) 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

$$\mu = E(Y) = 1 \cdot 11/36 + 2 \cdot 9/36 + 3 \cdot 7/36 + 4 \cdot 5/36 + 5 \cdot 3/36 + 6 \cdot 1/36 = \frac{11+18+21+20+15+6}{36} = 2.527.$$

3.2.6 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 3.14

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 $\sigma(X)$ of r.v. X is defined as $\sigma(X) = \sqrt{var(X)}$.

The variance exists if the second momdent $E(X^2)$ exists. The latter implies that the first moment $E(X)$ also exists.

Corollary 3.1

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.

$$\begin{aligned} E((X-E(X))^2) &= \sum_{x_i\in X}(x_i^2+E(X)*E(X)-2x_iE(X))f(x_i) \\ &= \sum_{x_i\in X}(x_i^2f(x_i))+\sum_{x_i\in X}(E(X)\cdot E(X))f(x_i)+\sum_{x_i\in X}(-2x_iE(X))f(x_i) \\ &= E(X^2)+E(X)\cdot E(X)\cdot\sum_{x_i\in X}f(x_i)-2E(X)\cdot\sum_{x_i\in X}x_if(x_i) \\ &= E(X^2)+E(X)\cdot E(X)-2E(X)\cdot E(X) \\ &= E(X^2)-E(X)\cdot E(X). \end{aligned}$$

For the continuous x a similar proof can be obtained. ■

Theorem 3.7

Let X be a discrete random variable with $E(X)=\mu$. Then

$$var(X)=E((X-\mu)^2)=\sum_{i=1}^{n}(x_i-\mu)^2P(X=x_i).$$

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 3.8

Let X and Y be two discrete random variables. The following apply.

$$E(X+Y)=E(X)+E(Y),$$

$$var(X+Y)=var(X)+var(Y), \quad \textit{for independent} \quad X,Y.$$

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 $z_k=x_i+y_j,\ldots,1\leq i,j\ldots$.

$$\begin{aligned} E(X+Y) &= \sum_k z_kP(X+Y=z_k) \\ &= \sum_i\sum_j(x_i+y_j)P(X=x_i,Y=y_j) \\ &= \sum_i x_i\sum_j y_jP(X=x_i,Y=y_j)+\sum_j y_j\sum_i x_iP(X=x_i,Y=y_j) \\ &= \sum_i x_iP(X=x_i)+\sum_j y_jP(Y=y_j) \\ &= E(X)+E(Y). \end{aligned}$$

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,

$$E(aX) = \sum_i a x_i P(aX = a x_i) = \sum_i a x_i P(X = x_i) = aE(X).$$

$$var(aX) = E((aX)^2) - E(aX)^2 = a^2 E(X^2) - a^2 E(X)^2 = a^2 var(X).$$

■

It becomes clear then that $var(aX + b) = a^2 var(X)$.

Corollary 3.2

For a r.v. X and any $a, b \in \mathbb{R}$, we have the following.

$$var(aX + b) = a^2 var(X).$$

3.2.7 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 3.15 **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 3.16

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,

$$P(X = x_i) = F_X(x_i) - F_X(x_{i-1}).$$

3.2.8 Probability generating function

Definition 3.17

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 3.9

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.

$$(g_X(1))'' + E(X) - E(X)^2 == E(X^2) - E(X) + E(X) - E(X)^2 = var(X).$$

■

3.2.9 Miscellanea

Theorem 3.10

Let X, Y be r.v. and $a, b \in \mathbb{R}$. Then

$$E(aX + bY) = aE(X) + bE(Y).$$

Proof.

$$E(aX + bY) = \sum_{s \in S} (aX + bY)(s)p(s) = a\sum_{s} X(s)p(s) + b\sum_{s} Y(s)p(s) = aE(X) + bE(Y).$$

■

Corollary 3.3

For X and Y r.v. as before,

$$E(X + Y) = E(X) + E(Y),$$

obtainable by setting $a = b = 1$.

The result below applies to any probability space.

Theorem 3.11

For any random variables $X_1, \ldots, X_n$ of a probability space, we have that

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i).$$

Proof. By induction on n. We alread proved it for $n = 2$, the base case. Inductive step follows trivially. ■

3.3 Continuous random variables

Definition 3.18 **Continuous r.v.**

A random variable is continuous if its domain is uncountably infinite.

3.3.1 Cumulative distribution function

Definition 3.19 **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 3.12

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.

3.3.2 Probability density function (p.d.f.)

Definition 3.20 **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

$$P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(t)dt.$$

3.3.3 Expectation of a continuous r.v.

Definition 3.21 **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 3.22 **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 3.23 **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

$$E(h(X,Y)) = \int_{-\infty}^{\infty} h(X,Y) f_{X,Y}(x,y)dx.$$

3.3.4 Independent random variables: continuous case

Definition 3.24 **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,

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

3.3.5 Variance of a continuous r.v.

Definition 3.25 **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 3.13

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 3.26 **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 3.14

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 3.15

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 independent $E(XY) = E(X)E(Y)$ and thus $cov(X,Y) = 0$. We the obtain the following.

Corollary 3.4

For r.v. X, Y that are independent we have the following.

$$var(X+Y) = var(X) + var(Y).$$

The results can be generalized for $n > 2$ variables. Proofs are by induction on n.

Definition 3.27 **Correlation**

The correlation of two random variable X,Y is given by the following expression.

$$Corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

Example 3.3.1

Let X be a r.v. such that $P(a \leq X \leq b) = 1$ in other words $a \leq X \leq b$. Show that

$$var(X) \leq \frac{(b-a)^2}{4}.$$

Proof. For any X, we have $var(X) \leq E((X-t)^2)$, for any t. (Easy to prove if one expands $var(X) = E((X-E(X))^2)$. Equality is for $t = E(X)$. Consider $t = (a+b/2$. This is the midpoint of the interval $[a,b]$. Therefore $|X-t| \leq (b-a)/2$. Therefore

$$var(X) \leq E((X-t)^2) \leq (b-a)^2/4.$$

■

3.3.6 Median and other quartiles

Definition 3.28

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

$$F_X(x_n) = n/100.$$

3.3.7 Logarithmic and other inequalities

The primary source for the listed inequalities is [1], [17], [2] , [11].

Inequality 3.3.1

For every $x > -1$ and $x \neq 0$ we have the following.

$$\frac{x}{1+x} < \ln(1+x) < x.$$

For every $x > 0$ we have the following.

$$\ln(x) \leq x - 1.$$

For every $x < 1$ and $x \neq 0$ we have the following.

$$x < -\ln(1-x) < \frac{x}{1-x}.$$

Inequality 3.3.2

For every $0 \leq \delta < 1$ we have the following.

$$(1-\delta)\ln(1-\delta) \geq -\delta + \delta^2/2.$$

Proof. 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

$$(1-x)\ln(1-x) \geq -x+x^2/2+x^2/6 \geq -x+x^2/2$$

■

Inequality 3.3.3
For every $0 \leq \delta \leq 1$ we have the following.

$$(1+\delta)\ln(1+\delta) \geq \delta+\delta^2/3.$$

Proof. For $0 \leq x \leq 1$,

$$\begin{aligned} \ln(1+x) &= x-x^2/2+x^3/3-x^4/4+\ldots \\ (1+x)\ln(1+x) &= x-x^2/2+x^3/3-x^4/4+\ldots \\ &+ \quad x^2-x^3/2+x^4/3-x^5/4+\ldots \\ (1+x)\ln(1+x) &\geq x+x^2/2-x^3/6 \end{aligned}$$

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. ■

Inequality 3.3.4
For every $2e-1 < \delta$ we have the following.

$$(1+\delta)\ln(1+\delta) \geq (2e)^{2e}.$$

Fact 3.1
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 3.2
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 $1-1/x < 1/e$.

Definition 3.29 **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 3.16

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. *$P(X = x) = F(x) - P(X < x)$.*

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.

Definition 3.30 **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.
We can also define a cumulative distribution function for a discrete r.v. X. Then $F(x) = \sum_{t=X(s)\leq x, s\in S} f(t)$.

3.3.8 Composition of random variables

Definition 3.31 **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.

$$\forall s\in S: f(X,Y,\ldots,Z)(s) = f(X(s),Y(s),\ldots,Z(s)).$$

3.4 Conditional expectation

Definition 3.32

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 3.33

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 3.34

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 3.35 **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 $f_{X,Y}(x,y) = f_{X|Y}(x|y)\cdot f_Y(y)$.

$$E(X|Y=y) = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx = \frac{1}{f_Y(y)}\int_{-\infty}^{\infty} xf_{X,Y}(x,y)dx.$$

Definition 3.36

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 3.37

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 3.4.1

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 3.38

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 3.39

If X,Y are random variables then

$$E(E(X|Y)) = E(X).$$

Theorem 3.17

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 $E(E(X|Y)|Y) = E(X|Y)$. ■

3.5 More on expectations

Theorem 3.18

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 3.19

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.

$$\begin{aligned}\int_{-\infty}^{0} P(X < t)dt &= \int_{-\infty}^{0} (\int_{-\infty}^{t} f_X(x)dx)dt \\ &= \int_{-\infty}^{0} (\int_x^{0} f_X(x)dt)dx \\ &= \int_{-\infty}^{0} (-x f_X(x))dx \\ &= \int_{-\infty}^{0} (-x f_X(x))dx.\end{aligned}$$

The combination yields the result. ■

Theorem 3.20

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

$$\begin{aligned} E(Z) &= \int_0^\infty P(Z > z)dz - \int_{-\infty}^0 P(Z < z)dz \\ &= \int_0^\infty \sum_i a_i P(X_i > z)dz - \int_{-\infty}^0 \sum_i a_i P(X_i < z)dz \\ &= \sum_i a_i E(X_i). \end{aligned}$$

■

Chapter 4

Standard distributions and random variables

4.1 Standard distributions

4.1.1 Uniform distribution

Definition 4.1

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.

$$f_X(x) = 1/(b-a), \quad a \leq x \leq b, \quad f_X(x) = 0 \quad \text{otherwise.}$$

Moreover,

$$F_X(x) = 0 \text{ if } x < a, \quad F_X(x) = 1 \text{ if } x > b, \quad \textit{and} \quad F_X(x) = \frac{x-a}{b-a} \text{ if } a \leq x \leq b.$$

Furthermore

$$E(X) = (a+b)/2,$$

and

$$var(X) = (b-a)^2/12.$$

4.1.2 Bernoulli distribution

Definition 4.2 **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 4.1

Let X be a random variable that follows a Bernoulli distribution with parameter p that is, $X \sim b(p)$. Sometimes we use $q = 1-p$. The probability of success and failure are $P(X=1) = p$ and $P(X=0) = 1-p = q$ respectively.

The expected value $E(X)$ of X and its variance $var(X) = \sigma^2$ are as follows.

$$E(X) = p, \qquad var(X) = pq = p(1-p).$$

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.

$$var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq.$$

■

4.1.3 Geometric distribution

Definition 4.3

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 4.2

Let X be a random variable that follows a geometric distribution with parameter p that is, $X \sim g(p)$. Sometimes we use $q = 1-p$. The probability of the first success after $k-1$ failures is $P(X=k) = (1-p)^{k-1} \cdot p$. The expected value $E(X)$ of X and its variance $var(X) = \sigma^2$ are as follows.

$$E(X) = 1/p, \qquad var(X) = (1-p)/p^2.$$

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 $\sum_{i=0}^{\infty}x^i=1/(1-x)$ 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 $x<1$.

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{\infty}k^2P(X=k)\\
&= \sum_{k=1}^{\infty}(k^2-k+k)(1-p)^{k-1}p\\
&= \sum_{k=1}^{\infty}k(k-1)(1-p)^{k-1}p+\sum_{k=1}^{\infty}k(1-p)^{k-1}p\\
&= p(1-p)\sum_{k=1}^{\infty}k(k-1)(1-p)^{k-2}+E(X)\\
&= p(1-p)(2/(1-(1-p))^3)+1/p\\
&= 1/p+2(1-p)/p^2
\end{aligned}$$

Then

$$var(X)=E(X^2)-(E(X))^2=1/p+2(1-p)/p^2-(1/p)^2=(1-p)/p^2.$$

■

4.1.4 Binomial distribution

Definition 4.4 **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 or trial or experiment of n independent Bernoulli trials each one $b(p)$.

Theorem 4.1

The probability of k successes in a binomial process $B(n,p)$ is denoted by $B(n,p;k)$ and given by

$$B(n,p;k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

Corollary 4.3

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 4.4

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 4.2 **Binomial Distribution $B(n,p)$**

For the binomial distribution, we have that $P(X=k)=B(n,p;k)=\binom{n}{k}p^kq^{n-k}$, where $q=1-p$. Then

$$\mu=E(X)=\sum_k kP(X=k)=np.$$

$$var(X) = E((X-\mu)^2) = npq.$$
$$\sigma(X) = \sigma_X = \sqrt{npq}.$$

Theorem 4.3

Let X be a random variable that follows a binomial distribution with parameters n and p that is, $X \sim B(n,p)$. Sometimes we use $q = 1-p$. The probability of k successes in n trials is B(n,p;k). The expected value $E(X)$ of X and its variance $var(X) = \sigma^2$ 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 the following.

$$var(X) = E(X^2) - E^2(X) = n^2p^2 + np(1-p) - (np)^2 = np(1-p) = npq.$$

Having provided those two detailed proofs, since X is the sum of n identical and independent Bernoulli random variables X_i, which implies $X = \sum_i X_i$, we have $E(X) = E(\sum_i X_i) = n \times E(X_i) = np$. By properties of variance $var(X) = \sum_v ar(X_i) = npq$, thus proving the result directly without mathematical manipulations. ■

Example 4.1.1

Let X_i, $i = 1, \ldots, n$ be n independent Bernoulli variables representing the outcomes of a sequence of n coin tosses with bias p, where $0 < p < 1$ that is, $X_i \sim b(p)$. Consider $A_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. The latter is the fraction of Heads in the n trials i.e. in space $X^n = (X_1, \ldots, X_n)$. Show the following.

$$E(A_n) = p, \quad var(A_n) = p(1-p)/n.$$

Proof. By way of the definition of $A_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ we have the following

$$E(A_n) = E(\frac{1}{n} \sum_i X_i) = \frac{1}{n} \sum_i E(X_i) = (1/n)np = p.$$

By the definition and properties of the variance, and the fact that X_I are independent we have the following.

$$var(A_n) = var(\frac{1}{n} \sum_i X_i) = \frac{1}{n^2} \sum_i var(X_i) = (1/n^2)n \cdot pq = pq/n.$$

■

4.1.5 Normal distribution

Definition 4.5

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 $Y = (X-\mu)/\sigma \sim N(0,1)$.

4.2 Exercises

Exercise 4.2.1 **Coin toss again**

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. $n! \approx \sqrt{2\pi n}(n/e)^n$

$$p = \frac{n!}{(n/2)!(n/2)!} \frac{1}{2^n} = \frac{\sqrt{2\pi n}(n/e)^n}{\sqrt{2\pi n/2}(n/2e)^{n/2}\sqrt{2\pi n/2}(n/2e)^{n/2}} \frac{1}{2^n} = \frac{\sqrt{2}}{\sqrt{\pi n}}$$

■

Chapter 5

Generating functions

5.1 Generating functions

In this and later sections we will interchangeably use the following notation for the exponential function: $\exp(Z)$ will denote e^Z.

5.1.1 Probability generating function

Definition 5.1

Let X be a discrete random variable. Then the probability generating function (p.g.f.) is denoted as

$$p_X(t) = E(t^X),$$

for $t \in \mathbb{R}$.

Theorem 5.1

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.

$$p_X^{(m)}(t) = m! P(X=m) + O(1)t(P(X=m+1) + \ldots).$$

Setting $t = 0$ the second summand vanishes and the general result is obtained. ■

5.1.2 Moments

Definition 5.2

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.

5.1.3 Moment generating function

Definition 5.3

The moment generating function $M_X(t)$ of a random variable X, with cumulative distribution function F_X is defined as follows, provided that the expectation $E(\exp(tX))$ 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.

$$M_X(t) = E(\exp(tX)) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx.$$

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 $m_n = E(X^n)$ of r.v. X.

Theorem 5.2

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 times 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 5.1

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 5.3

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

$$m_Y(t) = (M_X(t))^n.$$

5.1.4 Some examples

Example 5.1.1

The Uniform distribution $U(a,b)$ has

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Example 5.1.2

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 5.2

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. ■

Theorem 5.4

Let X_i be independent random variables, with MGF $M_{X_i}(t)$ respectively, $i = 1, \ldots, n$. Let $S_n = \sum_{i=1}^{n} X_i$. Then the following applies for the MGF of random variable S_n.

$$M_{S_n}(t) = M_{X_1}(t) \ldots M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Furthermore, if X_i are identically distributed to random variable X, then

$$M_{S_n}(t) = (M_X(t))^n.$$

Example 5.1.3
The binomial distribution $B(n,p)$ has

$$M_X(t) = (1-p+pe^t)^n.$$

Note that $M_X(0) = 1$. Moreover $M'_X(t) = n(1-p+pe^t)^{n-1}pe^t$, and $M'_X(0) = np = E(X)$. Similarly $M''_X(0) = E(X^2)$.

Example 5.1.4
The normal distribution $N(\mu, \sigma^2)$ has

$$M_X(t) = \exp(\mu t + 0.5\sigma^2 t^2).$$

Example 5.1.5
The Poisson distribution $Poisson(\lambda)$ has

$$M_X(t) = \exp(\lambda(e^t - 1)).$$

Example 5.1.6
The exponential distribution with (λ) has

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}.$$

Example 5.1.7
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 $e^t(1-p) < 1$. ■

5.2 Indicator random variable

Definition 5.4 **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 5.5

For any event A, we have

$$E(I_A] = P(A).$$

Proof.

$$E(I_A) = \sum_{s\in S} I_A(s)p(s) = \sum_{s\in A} I_A(s)p(s) + \sum_{s\notin A} I_A(s)p(s) = \sum_{s\in A} I_A(s)p(s) + \sum_{s\notin A} 0\cdot p(s) = P(A)$$

■

Theorem 5.6

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 5.2.1

The expected 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

$$E(A) = \sum_i E(A_i) = n\cdot 1/n = 1.$$

5.3 Characteristic function

Definition 5.5

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 5.7

Let X be any random variable with finite first k moments. Then the following apply.

$$\begin{aligned} c_X(0) &= 1 \\ c_X'(0) &= i \cdot E(X) \\ c_X'''(0) &= -i \cdot E(X^2) \\ &\cdots \\ c_X^{(m)}(0) &= i^m \cdot E(X^m). \end{aligned}$$

Part III

Inequalities in probability theory

Chapter 6

Probability inequalities

6.1 Probability inequalities

Definition 6.1 **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 $f''(x) \geq 0$.

6.1.1 Markov's inequality

Theorem 6.1 **Markov's inequality**

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. Note that we write $X(s)$ for an $s \in S$ rather that $x_i = X(s_i)$ for an $s_i \in S$ to keep things simple and to separate the s with $X(s) < a$ from $X(s) \geq a$.

$$\begin{aligned} E(X) &= \sum_{X(s)} X(s)p(s) \\ &= \sum_{0 \leq X(s) < a} X(s)p(s) + \sum_{X(s) \geq a} X(s)p(s) \\ &\geq \sum_{X(s) \geq a} X(s)p(s) \\ &\geq a \cdot \sum_{X(s) \geq a} p(s) \\ &= a \cdot P(X(s) \geq a) = a \cdot P(X \geq a). \end{aligned}$$

(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 $P(X > a) \leq E(X)/a$. An alternative unified proof is as follows.

6.1.2 Cauchy-Schwartz inequality

Theorem 6.2 **Cauchy-Schwartz inequality**

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 6.3 **Cauchy-Schwartz inequality**

Let X and Y be two random variables. with $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then

$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

Proof. Consider $E((X - tY)^2) \geq 0$. Then

$$E((X - tY)^2) = E(X^2) + t^2 E(Y^2) - 2tE(X)E(Y) \geq 0.$$

Choose $t = \frac{E(XY)}{E(Y^2)}$. Then we have

$$\begin{aligned} E(X^2) + \left(\frac{E(XY)}{E(Y^2)}\right)^2 E(Y^2) - 2\frac{E(XY)}{E(Y^2)}E(XY) &\geq 0 \\ E(X^2) + \frac{E^2(XY)}{E(Y^2)} - 2\frac{E^2(XY)}{E(Y^2)} &\geq 0 \\ E(X^2) - \frac{E^2(XY)}{E(Y^2)} &\geq 0 \\ E(X^2)E(Y^2) &\geq E^2(XY). \end{aligned}$$

The last inequality, after taking square roots concludes the proof. ■

6.1.3 Tsebyshev's or Tchebychef's inequality

From now use we use the more frequently spelled out Chebyshev.

Theorem 6.4 **Chebyshev's inequality**

Let X be a random variable 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.

$$P(|X-\mu| \geq \sigma t) = P(|X-\mu|^2 \geq \sigma^2 t^2) \leq E((X-\mu)^2)/\sigma^2 t^2 = \sigma^2/\sigma^2 t^2 = 1/t^2.$$

■

An alternative formulation follows.

Corollary 6.1 **Chebyshev variation 1**

An alternative form is the following bound.

$$P(|X-\mu| \geq t) \leq \frac{var(X)}{t^2} = \frac{\sigma^2}{t^2}.$$

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 6.2 **Chebyshev variation 2**

An alternative form is the following bound.

$$P(|X| \geq t) \leq \frac{E(X^2)}{t^2}.$$

One can obtain sharper bounds for small variance by considering $P(|X-\mu|^k \geq \sigma t)$, where $k > 2$.

Derivative results

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 6.3

Let S_n be a random variable that is defined as $S_n = X_1 + \ldots + X_n$, where X_i are independent random variables. Then $var(S_n) = \sum_i var(X_i)$. 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_n as defined, let $E(S_n) = \sum_i E(X_i)$.

$$\begin{aligned}
P(|S_n - E(S_n)| \geq nt) &\leq \frac{var(S_n)}{n^2t^2} \\
P(\frac{1}{n}|S_n - E(S_n)| \geq t) &\leq \frac{var(S_n)}{n^2t^2} \\
P(\frac{1}{n}|\sum_i X_i - E(\sum_i X_i)| \geq t) &\leq \frac{n\sigma^2}{n^2t^2} \\
P(\frac{1}{n}|\sum_i (X_i - E(X_i))| \geq t) &\leq \frac{\sigma^2}{nt^2}
\end{aligned}$$

■

Corollary 6.4

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 $S_n = \sum_i X_i$ as before. Show that $Pr(S_n \geq 3n/4) \leq 2/n$.

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 $S_n = \sum_i X_i$, we have $var(S_n) = n \cdot var(X_i) = n/4$, and $E(S_n) = n/2$. Applying Chebyshev's inequality,

$$P(S_n \geq 3n/4) = P(S_n - n/2 \geq n/4) = \frac{1}{2} \cdot P(|S_n - n/2| \geq n/4) = 0.5P(|S_n - E(S_n)| \geq n/4) \leq 0.5 \cdot var(S_n)/(n/4)^2 = 2/n.$$

■

Definition 6.2 **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.

Example 6.1.1

Let $X_1, \ldots X_n$ be bernoulli $b(p)$ random variables. Then $S_n = \sum_i X_i$ is such that $S_n \sim B(n, p)$. Then $E(S_n) = np = \mu$ and $var(S_n) = \sum_i var(X_i) = npq$ as noted earlier. Let us reformulate this as follows: let $nX \sim B(n, p)$ that is $S_n = nX$. For any $\varepsilon > 0$ we have the following using Chebyshev's inequality.

$$P(\frac{1}{n}|S_n - E(S_n)| > t) = P(|X - E(X)| > t) \leq pq/nt^2$$

Proof. By Chebyshev's inequality

$$P(|S_n - E(S_n)| > t) \leq var(X)/t^2$$

Starting with

$$P(\frac{1}{n}|S_n - E(S_n)| > t) = P(|S_n - E(S_n)| > n \cdot t)$$

and the latter by Chebyshev's inequality

$$P(|S_n - E(S_n)| > n \cdot t) \leq var(S_n)/n^2t^2$$

Therefore

$$P(\frac{1}{n}|S_n - E(S_n)| > t) \leq var(S_n)/n^2t^2 = \sum_i var(X_i)/n^2t^2 = npq/n^2t^2 = pq/nt^2.$$

Furthermore we note, $S_n/n = X$ and $E(S_n)/n = E(X)$ and the result follows. ■

6.1.4 Jensen's inequality

Theorem 6.5 **Jensen's inequality**

Let g be a convex function on $\mathbb{R}$ that is, $g : \mathbb{R} \mapsto \mathbb{R}$. Let X be a random variable, and $Y = g(X)$ another random variable defined. The the following applies

$$E(g(X)) \geq g(E(X)).$$

Proof.
Continuous case. Since g is convex for every t, graph of g lies above its tangent at t that is

$$g(x) \geq g(t) + S(x - t),$$

where S is the slope of the tangent on t. Then set $X = x$ and $t = E(X)$. We have the following, after taking expectations of both sides,

$$\begin{aligned}
g(x) &\geq g(t) + S(x - t) \Leftrightarrow \\
g(X) &\geq g(E(X)) + S(x - E(X)) \Leftrightarrow \\
E(g(X)) &\geq E(g(E(X))) + E(S(x - E(X))) \Leftrightarrow \\
E(g(X)) &\geq g(E(X)) + S(E(x) - E(X))) \Leftrightarrow \\
E(g(X)) &\geq g(E(X)).
\end{aligned}$$

Discrete case. For a discrete random variable the proof follows. Let $S = \{a_1, \ldots, a_n\}$ and let X be a random variable with $x_i = X(a_i)$. Then let $f(x_i) = P(X = x_i)$ and thus $\sum_i f(x_i) = 1$. Moreover, $E(X) = \sum_i x_i f(x_i)$. Let $g(x)$ be a convex function and thus for $c_1 + c_2 = 1$, we have $g(c_1x_1 + c_2x_2) \leq c_1g(x_1) + c_2g(x_2)$ and this generalizes for $n > 2$, so that for $c_1 + c_2 + \ldots + c_n = 1$, we have $g(c_1x_1 + \ldots c_nx_n) \leq c_1g(x_1) + \ldots + c_ng(x_n)$. We then obtain $E(g(X)) \geq g(E(X))$. The proof of the latter is by induction on $n \geq 2$. For the base case $n = 2$ using the convexity of g and the fact that $\sum_i f(x_i) = 1$, we have

$$\begin{aligned}
E(g(X)) &= f(x_1)g(x_1) + f(x_2)g(x_2) \\
&\geq g(f(x_1)x_1 + f(x_2)x_2) \\
&= g(E(X)),
\end{aligned}$$

as needed. For the inductive step, assuming that the result is true for $n-1$ and will be shown true for n, we have.

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{n} f(x_i)g(x_i) \\ &= f(x_1)g(x_1)+f(x_2)g(x_2)+\sum_{i=3}^{n} f(x_i)g(x_i) \\ &= (f(x_1)+f(x_2))\left(\frac{f(x_1)}{(f(x_1)+f(x_2)}g(x_1)+\frac{f(x_2)}{(f(x_1)+f(x_2)}g(x_2)\right)+\sum_{i=3}^{n} f(x_i)g(x_i). \end{aligned}$$

By the convexity of $g(x)$ the first two terms of the right-hand side get combined as follows

$$E(g(X)) \geq (f(x_1)+f(x_2))g\left(\frac{f(x_1)}{(f(x_1)+f(x_2)}x_1+\frac{f(x_2)}{(f(x_1)+f(x_2)}x_2\right)+\sum_{i=3}^{n} f(x_i)g(x_i). \tag{6.1}$$

On the right-hand size there are the $n-2$ terms of the sum (from $i=3$ through $i=n$) plus the newly generated one further on the left, a total of $n-1$. By the inductive step we have the following.

$$\begin{aligned} E(g(X)) &\geq (f(x_1)+f(x_2))g\left(\frac{f(x_1)}{(f(x_1)+f(x_2)}x_1+\frac{f(x_2)}{(f(x_1)+f(x_2)}x_2\right)+\sum_{i=3}^{n} f(x_i)g(x_i) \\ &\geq g\left((f(x_1)+f(x_2))\frac{f(x_1)}{(f(x_1)+f(x_2)}x_1+(f(x_1)+f(x_2))\frac{f(x_2)}{(f(x_1)+f(x_2)}x_2+\sum_{i=3}^{n} f(x_i)x_i\right) \\ &= g\left(f(x_1)x_1+f(x_2)x_2+\sum_{i=3}^{n} f(x_i)x_i\right) \\ &= g\left(\sum_{i=1}^{n} f(x_i)x_i\right) \\ &= g(E(X)), \end{aligned}$$

as needed.

■

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.

6.1.5 Law of large numbers

Definition 6.3 **Sample mean or sample average**

Let X be a r.v. with expectation (mean) $E(X)=\mu$ and standard deviation σ. Let an experiment is repeated n times with repetitions independent of each other. Let X_i be the random variable associated with the i-th repetition $X_i \sim X$. Then the sample mean or sample 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.

Definition 6.4 **Law of Large Numbers**

For any $r > 0$ the probablity that the sample mean $\bar{X}$ of n independent experiments has a value in the interval $[\mu - r, \mu + r]$ is to the limit for large n equal to 1.

$$lim_{n\to\infty} P(|X - \mu| \leq r) = 0$$

Theorem 6.6 **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

$$P(\Big|\frac{S_n}{n} - \mu\Big| \geq \varepsilon) \to 0,$$

as $n \to \infty$ and equivalently,

$$P(\Big|\frac{S_n}{n} - \mu\Big| < \varepsilon) \to 1,$$

as $n \to \infty$.

A proof follows 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.

$$\lim_{n\to\infty} P(|X - \mu| < \varepsilon) = 1.$$

6.1.6 Holder's inequality

Theorem 6.7 **Holder's inequality**

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.

$$E(|XY|) \leq \sqrt{(E(X^2)(E(Y^2)}.$$

6.1.7 Liapunov's inequality

Theorem 6.8 **Liapunov's inequality**

For any r.v. X, Y, and $a < b \in (1, \infty)$, the following applies.

$$E(|X|^a)^{1/a} \leq (E(|X|^b)^{1/b}.$$

6.1.8 Minkowski's inequality

Theorem 6.9 **Minkowski's inequality**

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 6.5

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. $E(W|Z) = aE(X|Z) + bE(Y|Z)$.

6.1.9 Tails of the binomial distribution

The following result in the form of Equations (6.3) and (6.4) appears in Feller [9] and in its complete form including Equation (6.2) as Theorem 1.1 in Bóllobás [3], where $B(n,p;m)$ is expanded and bounded,

Theorem 6.10 **Feller[9], Bóllobás[3]**

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 $m = \lceil upn \rceil$, $u > 1$,

$$P(S_n \geq m) \leq \frac{u}{u-1} B(n,p;m). \tag{6.2}$$

Furthermore,

$$P(S_n \geq m) \leq \frac{m(1-p)}{(m-np)^2}. \tag{6.3}$$

and

$$P(S_n \leq m) \leq \frac{(n-m)p}{(np-m)^2}. \tag{6.4}$$

The first form appears in [3]; the other two in [9].

Proof. Let $B(n,p;k) = \binom{n}{k} p^k (1-p)^{n-k}$. Consider

$$\frac{B(n,p;k)}{B(n,p;k-1} = 1 + \frac{(n+1)p-k}{k(1-p)}. \tag{6.5}$$

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 $(n+1)p = M$, for integer M, then $B(n,p;k) = B(n,p;k-1)$, since the fraction above is equal to 1. Then there are two maxima for k and $k-1$ or in other words for $M = (n+1)p$ and $M-1 = (n+1)p-1$. Otherwise there exists only one integer T such that $(n+1)p - 1 < T \leq (n+1)p$. The same conclusion can be drawn if

$$\frac{B(n,p;k+1)}{B(n,p;k} = \frac{(n-k)p}{(k+1)(1-p)}. \tag{6.6}$$

If $(n-k)p > (k+1)(1-p)$ or equivalently $k < (n+1)p - 1$ then the sequence is increasing, if $(n-k)p < (k+1)(1-p)$ or equivalently $k > (n+1)p - 1$ then the sequence is decreasing, and for $(n-k)p = (k+1)(1-p)$ or

$k=(n+1)p-1$, the sequence attains a maximum at two points $M-1=k=(n+1)p-1$ and $M=k+1=(n+1)p$. In the remainder we work with the second form. Note that $m=\lceil upn\rceil > (n+1)p$.

Therefore $B(n,p;m+1)/B(n,p;m)$ is a decreasing sequence. From Eq. (6.6) we have that

$$\frac{B(n,p;m+1)}{B(n,p;m} = \frac{(n-m)p}{(m+1)(1-p)}$$

Therefore $P(S_n \geq m)$ is a geometric series with largest term $B(n,p;m)$ and ratio at most $\frac{(n-m)p}{(m+1)(1-p)} \leq \frac{(n-m)p}{m(1-p)}$. Let us call the ratio r. Then

$$P(S_n \geq m) \leq \frac{1}{1-r} B(n,p;m). \tag{6.7}$$

For $r=\frac{(n-m)p}{m(1-p)}$ we have

$$\begin{aligned} r &= \frac{(n-m)p}{m(1-p)} \Rightarrow \\ \frac{1}{1-r} &= \frac{1}{1-\frac{(n-m)p}{m(1-p)}} \\ &= \frac{m(1-p)}{m-np} \end{aligned} \tag{6.8}$$

Therefore Eq. (6.7) by way of Eq. (6.8) becomes.

$$P(S_n \geq m) \leq \frac{1}{1-r} B(n,p;m) \leq B(n,p;m) \frac{m(1-p)}{m-np} \tag{6.9}$$

Between $M=(n+1)p$ and m there are at least $(m-np)$ terms that are at least $B(n,p;m)$. Then we have

$$\begin{aligned} (m-np)B(n,p;m) &\leq \sum_{i=M}^{m} B(n,p;i) \leq 1 \\ B(n,p;m) &\leq \frac{1}{(m-np)}. \end{aligned} \tag{6.10}$$

Therefore we conclude that Eq. (6.9) by way of Eq. (6.10) is as follows.

$$P(S_n \geq m) \leq B(n,p;m) \frac{m(1-p)}{m-np} \leq \frac{m(1-p)}{(m-np)^2}, \tag{6.11}$$

as needed in Eq. (6.3). We derive Eq. (6.4) by using the antisymmetry of the result of Eq. (6.3). In Eq. (6.4) we are interested in at most m successes therefore at least $n-m$ failures. Therefore we rewrite the just derived Eq. (6.3), where $q=1-p$ is replaced by p and m is replaced by $n-m$ and p is replaced by q i.e. $1-p$. We then obtain the following.

$$P(S_n \geq n-m) \leq \frac{(n-m)p}{((n-m)-n(1-p))^2} \leq \frac{(n-m)p}{(mp-n)^2}. \tag{6.12}$$

This is a bound for failures $\geq n-m$ that also translates to a bound for successes $\leq m$ and this is Eq. 6.4, as required. For Eq. 6.2 consider $r<1/u$ and thus in the derivation of Eq.6.8 $1/(1-r)<u/(u-1)$, as needed. ■

Chapter 7

Hoeffding's method

7.1 Hoeffding's inequality

We now present Theorem 1 and Theorem 2 of Hoeffding [13]. The work [13] deals with random variables that are not necessarily Bernoulli but are bounded e.g. $0 \leq X_i \leq 1$ or $a_i \leq X_i \leq b_i$. Theorem 1 [13] for the case of Bernoulli trials takes the form of Corollary 8.6 associated with a Chernoff bound.

That's why several times the terms Chernoff and Hoeffding refer to the same bound. Theorem 1 of [13] is the strongest among the Hoeffding bounds. The bounds are sufficient for large deviations. Otherwise one can use bounds available in [3] and [9]. Angluin-Valiant bounds [2] are weaker but more useful; the latter are or may be more useful for small p.

7.2 Derived right tails

Note that all random variables X in the two theorems that follow are to have finite first and second moments. This Theorem 1 of [13] follows.

Theorem 7.1

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.

$$\begin{aligned} P(\bar{X} - p \geq h)) &= P(S_n \geq (p+h)n) \\ &\leq \exp(-D((p+h)||p)n) \\ &\leq \left[\left(\frac{p}{p+h} \right)^{p+h} \left(\frac{1-p}{1-p-h} \right)^{1-p-h} \right]^n . \end{aligned} \tag{7.1}$$

Note that if $h > 1 - p$, then Eq. (7.1) remains true, and for $h = 1 - p$, the right-hand side can be replaced by the limit $h \to 1 - p$ which is p^n.

We provide a proof below for the Theorem following the techniques of the Chernoff's method also attributed to Cramér [7]. The theorem also appears as Theorem 5.1 in [15].

Proof. In order to prove Eq.(7.1) we perform the following transformation, using in the last step Markov's inequality, and before that the monotonically increasing function $X \mapsto e^{tX}$. Consider a $t > 0$ and the monotonically

increasing function $f(x) = e^{tx}$. Note that $0 < h < 1 - p$ below.

$$\begin{aligned} P(\bar{X} - p \geq h)) &= P(\frac{S_n}{n} - p \geq h) \\ &= P(S_n - E(S_n) \geq hn) \\ &= P(S_n - pn \geq hn) \\ &= P(S_n \geq (p+h)n) \qquad (7.2) \\ &= P(e^{tS_n} \geq e^{t(p+h)}) \\ &\leq \frac{E(e^{tS_n})}{e^{n(p+h)t}} \\ &= e^{-n(p+h)t} E(e^{tS_n}) \qquad (7.3) \end{aligned}$$

We then examine $E(e^{tS_n})$.

$$\begin{aligned} E(e^{tS_n}) &= E(e^{t\sum_i X_i}) \\ &= \prod_{i=1}^{n} E(e^{tX_i}) \qquad (7.4) \end{aligned}$$

By the convexity of function $f(x) = \exp(tx)$ we have the following: the line segment connecting the points $(a, f(a))$ and $(b, f(b))$ lies over $f(x)$. The equation of the line segment is as follows.

$$y - e^{ta} = \frac{e^{tb} - e^{ta}}{b-a}(x-a). \qquad (7.5)$$

Therefore,

$$y = e^{tb}\frac{x-a}{b-a} + e^{ta}\frac{b-x}{b-a}. \qquad (7.6)$$

Because of the convexity of f we also have the following.

$$e^{tx} \leq y = e^{tb}\frac{x-a}{b-a} + e^{ta}\frac{b-x}{b-a}. \qquad (7.7)$$

We now consider $E(e^{tX_i})$. We have the following.

$$E(e^{tX_i}) \leq e^{tb}\frac{E(X_i)-a}{b-a} + e^{ta}\frac{b-E(X_i)}{b-a}. \qquad (7.8)$$

Because $0 \leq X_i \leq 1$, $a = 0$ and $b = 1$, Eq.(7.8) can be simplified.

$$E(e^{tX_i}) \leq e^t E(X_i) + e^0(1 - E(X_i)) \leq e^t p_i + 1 - p_i, \qquad (7.9)$$

where $p_i = E(X_i)$. Then we plug Eq.(7.9) into Eq.(7.4). We derived the following.

$$\begin{aligned} E(e^{tS_n}) &= \prod_{i=1}^{n} E(e^{tX_i}) \\ &= \prod_{i=1}^{n} (e^t p_i + 1 - p_i) \qquad (7.10) \end{aligned}$$

Given that geometric means are at most their arithmetic means we have the following.

$$\left(\prod_{i=1}^{n}(e^t p_i + 1 - p_i)\right)^{1/n} \leq \sum_i \frac{(e^t p_i + 1 - p_i)}{n} \leq (e^t p + 1 - p). \qquad (7.11)$$

Therefore Eq.(7.11) into Eq.(7.10) yields the following.

$$\begin{aligned} E(e^{tS_n}) &= \prod_{i=1}^{n}(e^t p_i + 1 - p_i) \\ &\leq (e^t p + 1 - p)^n \end{aligned}$$

From Eq.(7.3) utilizing Eq.(7.12) we finally derive the following.

$$\begin{aligned} P(\frac{S_n}{n} - p \geq h) &\leq e^{-n(p+h)t} E(e^{tS_n}) \\ &\leq e^{-n(p+h)t}(e^t p + 1 - p)^n \qquad (7.12) \\ &\leq \exp(-n(p+h)t + n \cdot \ln(e^t p + 1 - p)) \\ &\leq \exp(g(t)) \qquad (7.13) \end{aligned}$$

We consider the monotonicity of function $g(t)$.

$$g^{'}(t) = -(p+h)n + npe^t / \ln(1 - p + pe^t). \qquad (7.14)$$

Setting $g^{'}(t) = 0$ and solving for t we have,

$$t_0 = \ln \frac{(p+h)(1-p)}{p(1-p-h)}, \qquad e^{t_0} = \frac{(p+h)(1-p)}{p(1-p-h)}. \qquad (7.15)$$

Moreover $(p+h)(1-p) \geq p(1-p-h)$ and thus $t_0 > 0$, since $h < 1-p$ and thus $1-p-h > 0$. Substituting Eq.(7.15) for t in $g(t)$ in Eq.(7.13) the result in the form of equation Eq.(7.1) follows. ■

Theorem 1 of [13] includes (weaker) upper bounds of Eq.(7.1) of the form $\exp(-nh^2 k(p))$, where

$$k(p) = \frac{1}{1-2p} \ln \frac{1-p}{p},$$

for $0 < p < 1/2$, and

$$k(p) = \frac{1}{2p(1-p)},$$

for $1/2 \leq p < 1$. The proof arguments are tedious and we refer to [13]. Furthermore, a weaker bound of the form $\exp(-2nh^2)$ can also be derived. This can be established also through Theorem 2 of [13] that is stated and proved below.

Theorem 2 of [13], utilizes the following result that is proven separately.

Proposition 7.1 **Hoeffding [13], 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 by Eq.(7.7), we have the following.

$$e^{tX} \leq \frac{x-a}{b-a} e^{tb} + \frac{b-x}{b-a} e^{ta}.$$

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 $ta = (x/(b-a)) \cdot a = (x/(b-a)) \cdot (p-1)(b-a) = x(p-1)$.

$$\begin{aligned} f(t) &= \ln(g(t)) \\ f(t) &= \ln\left(\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}\right) \\ &= \ln\left(e^{ta}\left(\frac{-a}{b-a}e^{t(b-a)} + \frac{b}{b-a}\right)\right) \\ &= \ln\left(e^{ta}\left((1-p)e^{t(b-a)} + p\right)\right) \\ &= \ln\left(e^{ta}\left((1-p)e^{x} + p\right)\right) \\ &= ta + \ln\left((1-p)e^{x} + p\right) \\ f(x) &= x(p-1) + \ln\left((1-p)e^{x} + p\right) \end{aligned} \tag{7.16}$$

Function $f(x)$ as indicated in Eq 7.16 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 (7.16) then yields, after recovering the original variable names,

$$\begin{aligned} E(e^{tX}) &\leq \exp(\ln(g(t))) \\ &\leq \exp(x^2/8) \\ &\leq \exp\left((b-a)t^2/8\right). \end{aligned}$$

■

Theorem 2 of [13] is stated and proved below. Whereas in Theorem 1 of [13] variables X_i were bounded in range by 0 and 1, the bounds next are variable, in the sense that $a_i \leq X_i \leq b_i$.

Theorem 7.2

Let X_i be an independent random variable, $i = 1, \ldots, n$. Let $S_n = \sum_i X_i$, $\bar{X} = S_n/n$, where $a_i \leq X_i \leq b_i$ and $p = E(\bar{X})$ and $E(S_n) = np$. Then for any h such that $0 < h < 1 - p$ we have the following.

$$P(\bar{X} - p \geq h)) = P(S_n \geq (p+h)n) \leq \exp\left(\frac{-2n^2h^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.17}$$

Proof. In order to prove Eq.(7.17) we work out similarly to the other bound of Eq.(7.1). Consider a $t > 0$ and the monotonically increasing and convex function $f(x) = e^{tx}$.

$$\begin{aligned} P(\bar{X} - p \geq h)) &= P(S_n - E(S_n) \geq hn) \\ &= P(e^{t(S_n - E(S_n))} \geq e^{thn}) \\ &\leq \frac{E(e^{t(S_n - E(S_n))})}{e^{htn}} \\ &\leq e^{-htn} E(e^{t(S_n - E(S_n))}) \\ &\leq \inf_{h>0} e^{-htn} E(e^{t(S_n - E(S_n))}) \end{aligned} \tag{7.18}$$

We then examine $E(e^{t(S_n - E(S_n))})$. Note that X_i are independent random variables for Eq.(7.19) to be valid.

$$\begin{aligned} E(e^{t(S_n - E(S_n))}) &= E(e^{t(\sum_i X_i - E(\sum_i X_i))}) \\ &= \prod_{i=1}^{n} E(e^{t(X_i - E(X_i))}) \end{aligned} \tag{7.19}$$

Since $E(X_i - E(X_i)) = 0$, the conditions of Proposition 7.1 are satisfied. Therefore

$$E(e^{t(X_i - E(X_i))} \leq \exp\left(\frac{t^2(b_i - a_i)^2}{8}\right). \tag{7.20}$$

Eq.(7.18) by way of Eq.(7.19) and Eq.(7.20) yields the following.

$$\begin{aligned} P(\bar{X} - p \geq h)) &\leq \inf_{h>0} e^{-htn} E(e^{t(S_n - E(S_n))}) \\ &\leq \inf_{h>0} e^{-htn} \prod_{i=1}^{n} E(e^{t(X_i - E(X_i))}) \\ &\leq \inf_{h>0} e^{-htn} \prod_{i=1}^{n} \exp\left(\frac{t^2(b_i - a_i)^2}{8}\right) \\ &\leq \inf_{h>0} \exp\left(-htn + t^2 \sum_{i=1}^{n} \frac{(b_i - a_i)^2}{8}\right) \qquad (7.21) \\ &\leq \inf_{h>0} \exp(g(t)) \qquad (7.22) \end{aligned}$$

Function $t(t)$ is a parabola. Its minimum is for

$$t_0 = \frac{4hn}{\sum_{i=1}^{n}(b_i - a_i)^2} \tag{7.23}$$

Substituting t_0 for t in Eq.(7.21) yields the following equation.

$$P(\bar{X} - p \geq h) \leq \inf_{h>0} \exp\left(-htn + t^2 \sum_{i=1}^{n} \frac{(b_i - a_i)^2}{8}\right) \tag{7.24}$$

$$\leq \exp\left(-\frac{2h^2n^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.25}$$

The proof is complete as Eq.(7.25) is Eq.(7.17). ■

By symmetry one can also prove the following e.g. by $Y_i = -X_i$.

Theorem 7.3

Let X_i be an independent random variable, $i = 1, \ldots, n$. Let $S_n = \sum_i X_i$, $\bar{X} = S_n/n$, where $a_i \leq X_i \leq b_i$ and $p = E(\bar{X})$ and $E(S_n) = np$. Then for any h such that $0 < h$ we have the following.

$$P(\bar{X} - p \leq -h)) = P(S_n \leq (p-h)n) \leq \exp\left(\frac{-2n^2h^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.26}$$

In [15], Theorem 7.1 appears as Theorem 5.1, and Theorem 7.2 and Theorem 7.3 appear as Theorem 5.7 there.

There are variants of Theorem 7.1 and Theorem 7.2. These include the following ones.

Corollary 7.1

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.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-2n^2\delta^2p^2}{n(b-a)^2}\right). \tag{7.27}$$

Proof. Set $h = \delta p$ in Theorem 7.1. Moreover $a_i = a$, $b_i = b$, $i = 1, \ldots, n$. ■

A similar corollary to Corollary 7.1 can be proven for $P(S_n \leq (1-\delta)pn)$, if one bounds the number of failures rather than successes with Theorem 7.2 or if $a_i = 0$ and $b_i = 1$, equivalently consider $Y_i = -X_i$. The upper bound would be identical then.

Corollary 7.2 appears in [15] as 5.3 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 7.2

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.

$$P(S_n \geq pn + h) \leq \exp\left(\frac{-2h^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.28}$$

Proof. Replace hn in Theorem 7.2 with h. ■

Moreover one can combine Corollary 7.2 and Corollary 7.6 to bound $P(|S_n - pn| \geq h)$.

The following is due to [2]. It also appears as Corollary 8.5 and Corollary 8.7

Theorem 7.4 **Angluin-Valiant[2]**

For every n, p, b with $0 \leq p \leq 1$ and $0 \leq b \leq 1$, we have the following.

$$\sum_{k=0}^{k=\lfloor (1-b)np \rfloor} B(n,p;k) \leq \exp(-b^2 np/2),$$

and

$$\sum_{k=\lceil (1+b)np \rceil}^{n} B(n,p;k) \leq \exp(-b^2 np/3).$$

For the case of a binomial distribution of Bernouli trials the following becomes available.

Corollary 7.3

Let X_i be an independent random variable, $i = 1,\ldots,n$, such that $X_i \sim b(p)$, where $0 < p < 1$. 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.

$$P(S_n \geq (1+\delta)pn) \quad \leq \quad \exp\left(-2n\delta^2 p^2\right). \tag{7.29}$$

Corollary 7.4

Let X_i be an independent random variable, $i = 1,\ldots,n$, such that $X_i \sim b(p)$, where $0 < p < 1$. 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.

$$P(S_n \geq pn + h) \quad \leq \quad \exp\left(\frac{-2h^2}{n}\right). \tag{7.30}$$

7.3 Derived left tails

Corollary 7.5

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 $0 < \delta < 1$ we have the following.

$$P(S_n \leq (1-\delta)pn) \quad \leq \quad \exp\left(\frac{-2n^2\delta^2 p^2}{n(b-a)^2}\right). \tag{7.31}$$

Proof. Bound the number of failures rather than successes with Theorem 7.2 or equivalently consider $Y_i = -X_i$. ■

Corollary 7.6 appears in [15] as 5.4 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 7.6

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 $0 < h < 1 - p$ we have the following.

$$P(S_n \leq pn - h) \quad \leq \quad \exp\left(\frac{-2h^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.32}$$

For the case of Bernoulli trials we have the following simplifications.

Corollary 7.7

Let X_i be an independent random variable, $i = 1, \ldots, n$, such that $X_i \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$ and $E(S_n) = np$. Then for any δ such that $0 < \delta < 1$ we have the following.

$$P(S_n \leq (1-\delta)pn) \quad \leq \quad \exp\left(-2n\delta^2 p^2\right). \tag{7.33}$$

Corollary 7.8

Let X_i be an independent random variable, $i = 1, \ldots, n$, such that $\leq X_i \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$ and $E(S_n) = np$. Then for any h such that $0 < h < 1 - p$ we have the following.

$$P(S_n \leq pn - h) \quad \leq \quad \exp\left(\frac{-2h^2}{n}\right). \tag{7.34}$$

7.4 Derived concentration bounds

The following Corollary combines Corollary 7.1 with Corollary 7.5.

Corollary 7.9

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 $0 < \delta < 1$ we have the following.

$$P(|S_n - np| \geq \delta pn) \quad \leq \quad 2\exp\left(\frac{-2n^2\delta^2 p^2}{n(b-a)^2}\right). \tag{7.35}$$

The following Corollary combines Corollary 7.2 with Corollary 7.6 instead.

Corollary 7.10

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 $0 < h < 1 - p$ we have the following.

$$P(|S_n - pn| \geq h) \quad \leq \quad 2\exp\left(\frac{-2h^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{7.36}$$

Chapter 8

Chernoff's method

8.1 Chernoff's inequality

In [4] bounds on the tails of a set of Bernoulli trials are discussed in the form of Theorem 1. Theorem 1 of [4] is restated below as Theorem 8.1 after some relevant definitions. One can extract a variety of bounds out of Theorem 8.1. Such bounds can lead to a sequence of lemmas such as Lemma 8.1, Lemma 8.2, Lemma 8.3, and Lemma 8.4 for the right tails, and Lemma 8.5, and Lemma 8.6, and Lemma 8.7 for the left tails, and their associated corollaries along with some obvious concentration bounds (left and right tail bounds).

Definition 8.1

Let X_i, $i = 1, \ldots, n$ be independent random variables identical in distribution to random variable X whose moment generating function is $M_X(t) = E(e^{tX})$, and its cumulative distribution function is $F_X(x) = P(X \leq x)$. Let $E(X) = p = \mu$ and define $S_n = \sum_{i=1}^{n} X_i$ and $E(S_n) = nE(X) = n\mu$, and

$$m(r) = \inf E(e^{t(X-r)}) = \inf e^{-rt} M_X(t) = \inf e^{-rt} E(e^{tX}).$$

The infimum is with respect to the t's values. $M_X(t)$ attains a minimum value $m(0)$.

Skipping some other details, the t value for the minimum is finite unless $P(X > 0) = 0$ or $P(X < 0) = 0$ and then $m(0) = P(X = 0)$. If $P(X \leq 0) > 0$ and $P(X \geq 0) > 0$ then $m(0) > 0$. The following is Theorem 1 of Chernoff [4]

Theorem 8.1 **Chernoff [4]**

If $E(X) < \infty$ and $r \geq E(X)$ then

$$P(S_n \geq nr) \leq (m(r))^n. \tag{8.1}$$

If $E(X) > -\infty$ and $r \leq E(X)$ then

$$P(S_n \leq nr) \leq (m(r))^n. \tag{8.2}$$

If $0 < u < m(r)$ and $E(X)$ might not exist,

$$\lim_{n\to\infty} \frac{(m(r) - u)^n}{P(S_n \leq nr)} = 0. \tag{8.3}$$

Proof. In order to prove Eq.(8.1) we perform the following transformation, using in the last step Markov's inequality, and before that the monotonically increasing function $X \mapsto e^{tX}$. Consider a $t > 0$ and the monotonically

increasing function $f(x) = e^{tx}$.

$$\begin{aligned} P(S_n \geq nr) &= P(tS_n \geq tnr) \\ &= P(e^{tS_n} \geq e^{nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{-nrt} E(e^{tS_n}) \qquad (8.4) \\ &= e^{-nrt} M_{S_n}(t). \qquad (8.5) \end{aligned}$$

We then examine $M_{S_n}(t) = E(e^{tS_n})$. Since $S_n = \sum_i X_i$, then

$$\begin{aligned} E(e^{tS_n}) &= E(e^{t\sum_i X_i}) \\ &= E(\prod_i e^{tX_i}) \\ &= \prod_i E(e^{tX_i}) \\ &= \prod_i M_{X_i}(t) \\ &= (M_X(t))^n. \qquad (8.6) \end{aligned}$$

From Eq. (8.5) by way of Eq. (8.6) we obtain the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt} E(e^{tS_n}) \\ &\leq e^{-nrt} (M_X(t))^n \\ &= \left(e^{-rt} M_X(t)\right)^n \\ &= \inf_{t>0} \left(e^{-rt} M_X(t)\right)^n \\ &= \inf_{t>0} (m(r))^n. \qquad (8.7) \end{aligned}$$

This completes the proof of Eq.(8.1).
The proof of Eq.(8.2) is similar. Consider a $t < 0$ now.

$$\begin{aligned} P(S_n \leq nr) &= P(tS_n \geq tnr) \\ &= P(e^{tS_n} \geq e^{nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{-nrt} E(e^{tS_n}) \\ &= e^{-nrt} M_{S_n}(t). \qquad (8.8) \end{aligned}$$

The rest is identical to the previous case. ■

In most of the discussion to follow we will assume that X_i and X follow a Bernoulli distribution i.e. $X_i \sim b(p)$ and $X \sim b(p)$ and thus $E(X_i) = E(X) = p$, where $0 < p < 1$. Then, $S_n \sim B(n, p)$. Therefore we have the following.

$$M_X(t) = E(e^{tX}) = e^t \cdot p + 1 \cdot (1-p) = 1 + p(e^t - 1) \leq e^{pe^t - p}.$$

The last part is because for all x, we have $e^x \geq 1 + x$.

Definition 8.2 **Kullback-Leibler**

The Kullback-Leibler divergence D_{KL} or just simply D is defined as follows for two distributions of n elements $P = \cup_i \{p_i\}$, $p_i \geq 0$, $Q = \cup_i \{q_i\}$, $q_i \geq 0$, $i = 1, \ldots, n$ such that $\sum_i p_i = \sum_i q_i = 1$.

$$D_{KL}(P||Q) = D(P||Q) = \sum_i p_i \ln \frac{p_i}{q_i}. \tag{8.9}$$

8.2 Derived right tails

We derive the first Chernoff bound for a Binomial r.v. S_n which is the sum of n Bernoulli random variables, using Eq.(8.1).

Lemma 8.1

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.

$$P(S_n \geq rn) \leq \exp(-D(r||p)n) = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r} \right]^n. \tag{8.10}$$

Sometimes $r = p + t$ and the bound $p < r < 1$ becomes $p < p + t < 1$ or equivalently $0 < t < 1 - p$. We report this variant in Corollary 8.6.

Proof. (Of Eq. (8.10))
By Eq. (8.7) we have the following considering that $M_X(t) = E(e^{tX}) = 1 + p(e^t - 1) \leq e^{pe^t - p}$.

$$\begin{aligned} P(S_n \geq nr) &\leq \inf_{t>0} \left(e^{-rt} M_X(t)\right)^n \\ &\leq \inf_{t>0} \left(e^{-rt}(1 + p(e^t - 1))\right)^n \\ &\leq \inf_{t>0} \left(\frac{(1 + p(e^t - 1))}{e^{rt}}\right)^n \qquad (8.11) \\ &\leq \inf_{t>0} (\exp(f(t)))^n \qquad (8.12) \end{aligned}$$

As indicated by Eq. (8.12), $f(t) = \ln((1 + p(e^t - 1)) - rt$. Consider $f'(t) = pe^t/(1 + pe^t - p) - r$. Equating to zero $f'(t) = 0$ and solving for t we obtain $e^t = (1-p)r/(p(1-r))$. Continuing with the second derivative we find $f''(t) = p(1-p)e^t/(1 + pe^t - p)^2 > 0$. Therefore $f(t)$ has a minimum for $e^t = (1-p)r/(p(1-r))$. We then continue with Eq. (8.11) as follows.

$$P(S_n \geq nr) \leq \inf_{t>0} \left(\frac{(1 + p(e^t - 1))}{e^{rt}}\right)^n \tag{8.13}$$

The denominator $\exp(rt)$ for $e^t = (1-p)r/(p(1-r))$ is as follows.

$$\exp(rt) = ((1-p)r/(p(1-r)))^r = \frac{(1-p)^r r^r}{p^r (1-r)^r}. \tag{8.14}$$

Likewise the numerator is as follows.

$$(1 + p(e^t - 1)) = \frac{(1-p)rp}{p(1-r)} + 1 - p = \frac{1-p}{1-r} \tag{8.15}$$

Therefore we have the following for the quantity below.

$$\frac{(1+p(e^t-1))}{e^{rt}} = \frac{1-p}{1-r}\cdot\frac{p^r(1-r)^r}{(1-p)^r r^r} = \left[\left(\frac{p}{r}\right)^r\left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{8.16}$$

The proof is completed. ■

We generate one more bound from Eq.(8.1). The bound is stronger than the Corollaries to follow that bring it into an easier to deal with form.

Lemma 8.2

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 $r = (1+\delta)p$, $\delta > 0$.

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} \tag{8.17}$$

Proof. By Eq. (8.7) we use $M_X(t) = E(e^{tX}) = 1 + p(e^t - 1) \leq e^{pe^t - p}$, to obtain the following.

$$\begin{aligned} P(S_n \geq nr) &\leq \inf_{t>0}\left(e^{-rt}(1+p(e^t-1))\right)^n \\ &\leq \inf_{t>0}\left(e^{-rt}e^{pe^t-p}\right)^n \qquad (8.18) \\ &\leq \inf_{t>0}\left(\exp(pe^t - p - rt)\right)^n \\ &\leq \inf_{t>0}(\exp(f(t)))^n \qquad (8.19) \end{aligned}$$

The second to last part is because for all x, we have $e^x \geq 1 + x$. By Eq. (8.18) and Eq. (8.19) we have $f(t) = pe^t - p - rt$. Since $f'(t) = pe^t - r$, setting $f'(t) = 0$ we obtain $t = \ln(r/p)$. Moreover, $f''(t) = pe^t$ is equal to $f''(\ln(r/p)) = r > 0$. Therefore $f(t)$ has a minimum at $t = \ln(r/p)$. Substituting this value for t in Eq.(8.19) the following is obtained.

$$\begin{aligned} P(S_n \geq nr) &\leq \inf_{t>0}\left(\exp(pe^t - p - rt)\right)^n \\ &\leq (\exp(p(r/p) - p - r\ln(r/p))^n \\ &\leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n \qquad (8.20) \end{aligned}$$

Finally we substitute $r = (1+\delta)p$ in Eq. (8.20) to obtain our result

$$\begin{aligned} P(S_n \geq nr) &\leq \left(\frac{e^{(1+\delta)p}\cdot p^{(1+\delta)p}}{e^p \cdot ((1+\delta)p)^{(1+\delta)p}}\right)^n \\ &\leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} \qquad (8.21) \end{aligned}$$

■

A simpler proof of Eq. (8.10) for binomial random variables is found in [5] and the proof is presented below.

Lemma 8.3 **[5]**

Let X_i be independent Bernoulli 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.

$$P(S_n \geq rn) \leq \exp\left(-D(r||p)n\right) = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{8.22}$$

Proof. Let $B(n,p,k) = \sum_{i=k}^{n} B(n,p;i)$. For any $x \geq 1$, we have

$$\begin{aligned} B(n,p,k) &\leq \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^{i-k} \\ &\leq x^{-k} \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^i \\ &\leq x^{-k} \sum_{i=0}^{n} \binom{n}{i} (px)^i (1-p)^{n-i} \\ &\leq x^{-k}(1+(x-1)p)^n. \end{aligned} \tag{8.23}$$

For $r > p$ consider $x = (1-p)r/(p(1-r))$ and substitute for the x of Eq. (8.23). We obtain the following.

$$\begin{aligned} B(n,p,k) &\leq x^{-k}(1+(x-1)p)^n \\ &\leq \left(\frac{(1-p)r}{p(1-r)}\right)^{-rn} \cdot \left(1+\left(\frac{(1-p)r}{p(1-r)}-1\right)p\right)^n \end{aligned} \tag{8.24}$$

The result then follows. ■

We present an alternative Chernoff bound formulation which follows the simple case of Hoeffding bounds, where $a_i = a = 0$ and $b_i = b = 1$. Naturally it applies to Bernoulli or binary random variables (e.g. Rademacher).

Lemma 8.4

Let X_i be independent Bernoulli 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.

$$P(S_n \geq rn) \leq \exp\left(-2n(r-p)^2\right). \tag{8.25}$$

Proof. The proof follows the steps of the proof of Theorem 8.1 to Eq.(8.4).

$$P(S_n \geq nr) = P(tS_n \geq tnr) \leq \frac{E(e^{tS_n})}{e^{nrt}} \leq e^{-nrt} E(e^{tS_n}) \tag{8.26}$$

We then proceed differently as follows.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt} E(e^{tS_n}) \\ &\leq e^{-nrt} e^{npt} E(e^{t(S_n - np)}) \\ &\leq e^{-nrt+npt} E(\prod_i e^{t(X_i - p)}) \\ &\leq e^{-nrt+npt} \prod_i E(e^{t(X_i - p)}) \end{aligned} \tag{8.27}$$

We note that r.v. $X_i - p$ is bounded and $E(X_i - p) = E(X_i) - p = p - p = 0$. Then, Proposition 7.1 that is being utilized in Hoeffding bounds can be used to show the following.

$$E(e^{t(X_i-p)}) \leq \exp\left(\frac{t^2}{8}\right). \tag{8.28}$$

Eq.(8.27) by way of Eq.(8.28) yields the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt+npt} \prod_i E(e^{t(X_i-p)}) \\ &\leq e^{-nrt+npt} \prod_i \exp\left(\frac{t^2}{8}\right) \\ &\leq e^{-nrt+npt} \exp\left(n\frac{t^2}{8}\right) \\ &\leq e^{-nrt+npt+\frac{nt^2}{8}}. \end{aligned} \tag{8.29}$$

Consider the exponent of Eq.(8.29): $f(t) = -nrt + npt + nt^2/8$. Its first derivative is $f^{'}(t) = -nr + np + t/4$. Equating it to zero and solving for t we have that $f^{'}(t) = 0 = -nr + np + t/4$ yields

$$t = 4(r - p). \tag{8.30}$$

Given that $f^{''}(t) = 1/4 > 0$, we have a minimum at $t = 4(r-p)$ for $f(t)$. Therefore Eq.(8.29) by way of Eq.(8.30) yields the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt+npt+\frac{nt^2}{8}} \\ &\leq e^{-nr\cdot 4(r-p)+np\cdot 4(r-p)+\frac{n\cdot 16(r-p)^2}{8}} \\ &\leq e^{-2n(r-p)^2}. \end{aligned} \tag{8.31}$$

Eq.(8.31) is Eq.(8.25) as needed. ■

The following corollary is more widely known than Lemma 8.4.

Corollary 8.1

Let X_i be independent Bernoulli 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.

$$P(S_n - E(S_n) \geq rn) \leq \exp\left(-2nr^2\right). \tag{8.32}$$

Proof. In Lemma 8.4 substitute $r + p$ for r. Then $P(S_n \geq (r+p)n) = P(S_n \geq rn + pn) = P(S_n - E(S_n) \geq rn)$ and then substitute $r + p$ for r in Eq.(8.25) to obtain Eq.(8.32). ■

The following Corollary can be obtained from Lemma 8.2. It provides a more tangible upper bound than the generic one of the Lemma.

Corollary 8.2

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,

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq 2^{-(1+\delta)pn}. \tag{8.33}$$

Proof. From Lemma 8.2, we have

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq \frac{e^{1+\delta}}{(2e)^{(1+\delta)}} \leq \frac{e^{1+\delta}}{(2e)^{1+\delta}} \leq 2^{-(1+\delta)}.$$

The result follows then. ■

The following Corollary is also obtained from Lemma 8.2. Note that $\delta > 0$ is better than the δ used in Corollary 8.5 that follows it.

Corollary 8.3

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 $\delta > 0$ we have the following.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{2+\delta} \cdot pn\right). \tag{8.34}$$

Proof. We would like to upper bound the bound of Eq.(8.17) of Lemma 8.2, as follows.

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq \exp\left(\frac{-\delta^2}{2+\delta}\right)$$

Consider function $f(t)$ defined as follows.

$$f(t) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{2+\delta}.$$

Note that for any $\delta > 0$ we have the following.

$$\ln(1+\delta) \geq \frac{2\delta}{2+\delta}.$$

Therefore we have the following result

$$f(t) \leq \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{2+\delta} \leq \frac{\delta(2+\delta) - (1+\delta)(2\delta) + \delta^2}{2+\delta} = 0$$

The Corollary then follows. ■

A small improvement has been proposed by McDiarmid in the following form.

Corollary 8.4 **[16]**

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 $\delta > 0$ we have the following.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{2+2\cdot\delta/3} \cdot pn\right). \tag{8.35}$$

Proof. The proof utilizes the following inequality for all $x > 0$.

$$(1+x)\ln(1+x) - x \geq \frac{x^2}{2+(2/3)x}$$

■

The following Corollary can also be obtained from Lemma 8.2. It is similar to the one in [2] for the binomial case.

Corollary 8.5

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.

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{8.36}$$

Proof. By inequality (9.4.8) we have for $0 < \delta < 1$, $(1+\delta)\ln(1+\delta) \geq \delta + \delta^2/3$. This results to

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq \frac{e^{\delta}}{e^{\delta+\delta^2/3}} \leq e^{-\frac{\delta^2}{3}}.$$

The result follows. Note that as proved, $\delta < 1$. However, we can improve the upper bound $\delta \leq 1$ by a more tedious approach. We show it below. Consider as before in Corollary 8.3, function $f(t)$ (t substitutes for δ) defined as follows.

$$f(t) = t - (1+t)\ln(1+t) + \frac{t^2}{3}.$$

We would like to show $f(t) \geq 0$. We first calculate its first derivative.

$$f^{'}(t) = 2t/3 - \ln(1+t).$$

We note that $f^{'}(0) = 0$ and $f^{'}(1) < 0$. In order to study the monotonicity of $f^{'}(t)$ we go on calculating the second derivative.

$$f^{''}(t) = 2/3 - 1/(1+t).$$

We note that $f^{''}(0) < 0$ for $t \leq 1/2$ and $f^{''}(1) > 0$ for $t > 1/2$. This means that $f^{'}(t)$ is monotonically decreasing for $t \leq 1/2$ and given $f^{'}(0) = 0$, negative for $t \leq 1/2$, and monotonically increasing and since $f^{'}(1) < 0$ also negative for $1 > t > 1/2$. One can also separately confirm that $f(1/2) < 0$. Thus $f(\delta) \leq 0$ for all $0 < \delta \leq 1$. The $\delta > 0$ is needed since $r > p$. ■

We report below a corollary variant of Lemma 8.1. This is Corollary 8.6.

Corollary 8.6

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 t such that $0 < t < 1 - p$ we have the following.

$$P(S_n \geq (p+t)n) \leq \exp(-D((p+t)||p)n) = \left[\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{1-p}{1-p-t}\right)^{1-p-t}\right]^n. \tag{8.37}$$

8.3 Derived left tails

We proceed to deriving a Lemma identical to Lemma 8.1 for the left tails. This is stated next.

Lemma 8.5

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 $0 < r < p$ we have the following.

$$P(S_n \leq rn) \leq \exp(-D(r||p)n) = \left[\left(\frac{p}{r}\right)^{r}\left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{8.38}$$

Sometimes $r = p - t$, and the bound $0 < r < p$ becomes $0 < p - t < p$ or equivalently $0 < t < p$.

Proof. (Of Eq. (8.38))
Method 1. A simple argument works as follows: the upper bound on the number of successes generates a corresponding lower bound on the number of failures. Thus for $F_i = 1 - X_i$, we have $\sum F_i = n - \sum_i X_i$ or equivalently $Y_n = n - S_n$. Therefore

$$P(S_n \leq rn) = P(n - S_n \geq n(1-r)) = P(Y_n \geq n(1-r))$$

The latter bound by Lemma 8.1 is bounded above by Eq.(8.10) adjusting it with r replaced by the $1-r$ of $Y_n \geq n(1-r)$ and p by $1-p$ to account for the failures not the successes of random variable Y_n. The end result is that

$$P(S_n \leq rn) \leq \exp(-D((1-r)||(1-p))n).$$

However $D((1-r)||(1-p)) = D(r||p)$ and therefore

$$P(S_n \leq rn) \leq \exp(-D((1-r)||(1-p))n) = D(r||p).$$

Method 2. Now let us reprove it following the method of the proof of Lemma 8.1. Consider $t > 0$, and apply the Chernoff trick and finally use Markov's inequality as we did earlier.

$$\begin{aligned} P(S_n \leq nr) &= P(tS_n \leq tnr) \\ &= P(-tS_n \geq -tnr) \\ &= P(e^{-tS_n} \geq e^{-nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{nrt} E(e^{-tS_n}) \\ &\leq e^{nrt} (E(e^{-Xt}))^n \end{aligned} \tag{8.39}$$

We calculate $E(e^{-tX}) = 1 + p(e^{-t} - 1) \leq e^{pe^{-t} - p}$. The rest of the calculation are similarly to the ones before

$$\begin{aligned} P(S_n \leq nr) &\leq \inf_{t>0} \left(e^{rt}(1 + p(e^{-t} - 1))\right)^n \\ &\leq \inf_{t>0} (\exp(f(t)))^n. \end{aligned} \tag{8.40}$$

As indicated by Eq. (8.40), $f(t) = \ln(1 + p(e^{-t} - 1)) + rt$. Consider $f'(t) = -pe^{-t}/(1 + p^{-t} - p) + r$. Equating to zero $f'(t) = 0$ and solving for t we obtain $e^{-t} = (1-p)r/(p(1-r))$. Continuing with the second derivative we find $f''(t) = p(1-p)e^{-t}/(1 + pe^{-t} - p)^2 > 0$. Therefore $f(t)$ has a minimum for $e^{-t} = (1-p)r/(p(1-r))$. We then continue with Eq. (8.40) as follows.

$$P(S_n \leq nr) \leq \inf_{t>0} \left(e^{rt}(1 + p(e^{-t} - 1))\right)^n \tag{8.41}$$

The term $\exp(rt)$ for $e^{-t} = (1-p)r/(p(1-r))$ is as follows.

$$\exp(rt) = \frac{p^r (1-r)^r}{(1-p)^r r^r}. \tag{8.42}$$

Likewise the other term is as follows.

$$(1 + p(e^{-t} - 1)) = \frac{p(1-r) - p^2(1-r) + p(1-p)r}{p(1-r)} = \frac{1-p}{1-r} \tag{8.43}$$

Therefore we have the following for the quatinty below.

$$e^{rt} \cdot (1 + p(e^{-t} - 1)) = \frac{1-p}{1-r} \cdot \frac{p^r(1-r)^r}{(1-p)^r r^r} = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r} \right]^n. \tag{8.44}$$

■

We generate one more bound below.

Lemma 8.6

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 $0 < r < p$ or, equivalently, for any δ such that $0 < \delta < 1$ we have the following.

$$P(S_n \leq rn) = P(S_n \leq (1-\delta)pn) \leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n = \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{pn} \tag{8.45}$$

Proof. We calculated earlier $E(e^{-tX}) = 1 + p(e^{-t} - 1) \leq e^{pe^{-t}-p}$. By way of Eq.(8.40) we have the following.

$$\begin{aligned} P(S_n \leq nr) &\leq \inf_{t>0} \left(e^{rt}(1 + p(e^{-t} - 1))\right)^n \\ &\leq \inf_{t>0} \left(\exp\left(pe^{-t} - p + rt\right)\right)^n \qquad (8.46) \\ &\leq \inf_{t>0} \left(\exp\left(f(t)\right)\right)^n \qquad (8.47) \end{aligned}$$

By Eq. (8.46) and Eq. (8.47) we have $f(t) = pe^{-t} - p + rt$. Since $f'(t) = r - pe^{-t}$, setting $f'(t) = 0$ we obtain $t = \ln(p/r)$. Moreover, $f''(t) = pe^t$ is equal to $f''(\ln(p/r)) = r > 0$. Therefore $f(t)$ has a minimum at $t = \ln(p/r)$. Substituting this value for t in Eq.(8.46) the following is obtained.

$$\begin{aligned} P(S_n \leq nr) &\leq \inf_{t>0} \left(\exp\left(pe^{-t} - p + rt\right)\right)^n \\ &\leq \left(\exp\left(pr/p - p + r\ln(p/r)\right)\right)^n \\ &\leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n \qquad (8.48) \end{aligned}$$

Finally we substitute $r = (1-\delta)p$ in Eq. (8.48) to obtain Eq.(8.45).

$$\begin{aligned} P(S_n \leq nr) &\leq \left(\frac{e^{(1-\delta)p} \cdot p^{(1-\delta)p}}{e^p \cdot ((1-\delta)p)^{(1-\delta)p}}\right)^n \\ &\leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{pn} \qquad (8.49) \end{aligned}$$

■

There is a weaker but more easier to deal bound for small p. This is shown next. It is similar to the one in [2] for the binomial case.

Corollary 8.7

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.

$$P(S_n \leq (1-\delta)pn) \leq \exp\left(\frac{-\delta^2}{2} \cdot pn\right). \tag{8.50}$$

Proof. By inequality (9.4.7) we have for every $0 < \delta < 1$, $(1-\delta)\ln(1-\delta) \geq -\delta - \delta^2/2$. This results to

$$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq \frac{e^{-\delta}}{e^{-\delta-\delta^2/2}} \leq e^{-\frac{\delta^2}{2}}.$$

The result follows. ■

The symmetric case for the left tails to Lemma 8.4 is stated below.

Lemma 8.7

Let X_i be independent Bernoulli 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 $0 < r < p$ we have the following.

$$P(S_n \leq rn) \leq \exp\left(-2n(r-p)^2\right). \tag{8.51}$$

Proof. The proof is by symmetry to Lemma 8.4 for $Y_i = 1 - X_i$ and $\sum_i Y_i = n - S_n$ instead. ■

The following corollary is then evident.

Corollary 8.8

Let X_i be independent Bernoulli 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 $0 < r < p$ we have the following.

$$P(S_n - E(S_n) \leq rn) \leq \exp\left(-2nr^2\right). \tag{8.52}$$

8.4 Derived concentration bounds

Finally the following Corollary can also be obtained from Corollary 8.5 and Corollary 8.7.

Corollary 8.9

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,

$$P(|S_n - np| \geq \delta pn) \leq 2 \cdot \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{8.53}$$

Proof. We have the following

$$P(|S_n - np| \geq \delta pn) = P(S_n - np \geq \delta pn) + P(S_n - np \leq -\delta pn) = P(S_n \geq (1+\delta)pn) + P(S_n \leq (1-\delta)pn)$$

By Corollary (8.5) we bound $P(S_n \geq (1+\delta)pn)$. By Corollary 8.7 we bound $P(S_n \leq (1-\delta)pn)$. Simple manipulations show $\exp\left(\frac{-\delta^2}{2}\right) < \exp\left(\frac{-\delta^2}{3}\right)$. The result then follows. ■

The following is a direct consequence of Corollary 8.1 and Corollary 8.8.

Corollary 8.10

Let X_i be independent Bernoulli 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 $0 < r < p$ we have the following.

$$P(|S_n - E(S_n)| \geq rn) \leq 2 \cdot \exp\left(-2n(r-p)^2\right). \tag{8.54}$$

Chapter 9

Combinatorial inequalities

9.1 Combinatorial inequalities

Inequality 9.1.1
For any integer $n >$ we have the following.

$$n! \leq n^n, \quad \text{and} \quad n! \geq (n/2)^{n/2}.$$

We also have the following.

$$n! \leq \left(\frac{n}{2}\right)^{\frac{n}{2}} n^{n/2} \quad \text{and} \quad n! \geq n^{\frac{n}{2}-\frac{\sqrt{n}}{2}}.$$

Proof.
Starting with $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$, we have that

$$\begin{aligned} n! &= 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n \\ &\leq n \cdot n \cdot \ldots \cdot n \cdot n \\ &\leq n^n. \end{aligned}$$

Likewise $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$ leads to

$$\begin{aligned} n! &= 1 \cdot 2 \cdot \ldots \cdot n/2 \cdot \ldots (n-1) \cdot n \\ &\geq 1 \cdot 1 \cdot \ldots \cdot n/2 \cdot \ldots n/2 \cdot n/2 \\ &\geq (n/2)^{(n/2)}. \end{aligned}$$

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. ■

Stirling's approximation for the factorial.

Inequality 9.1.2
For any integer $n > 0$ we have the following.

$$(n/e)^n \leq n! \leq en(n/e)^n.$$

Inequality 9.1.3
For any integer $n > 0$ and $1 \leq k \leq n$ we have the following.

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k.$$

Proof.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \\
&= \frac{n^k}{k!} \quad \text{Stirling's approximation} \\
&\leq \frac{n^k}{(k/e)^k} \\
&= \left(\frac{en}{k}\right)^k
\end{aligned}$$

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 $k \leq n$. ■

Inequality 9.1.4
If $k = o(\sqrt{n})$ then

$$\binom{n}{k} = (1+o(1))\frac{n^k}{k!}$$

Proof.

$$\binom{n}{k} = \frac{\prod_{i=0}^{k-1}(n-i)}{k!}$$

Then

$$n^k \geq \prod_{i=0}^{k-1}(n-i) \geq (n-k)^k \geq n^k(1-\frac{k}{n})^k \geq n^k(1-\frac{k^2}{n}) \geq n^k(1-o(1)).$$

Furthermore by Stirling's approximation

$$k! = (1+o(1)\sqrt{2\pi k}(k/e)^k$$

we obtain

$$\binom{n}{k} = (1+o(1))\frac{1}{\sqrt{2\pi k}}\left(\frac{ne}{k}\right)^k.$$

■

Inequality 9.1.5
If $k = \omega(1)$ and $k = o(n)$ then

$$\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k} \Rightarrow \lg\binom{n}{k} = (1+o(1))k\lg\frac{n}{k}.$$

Proof.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Then by Stirling's approximation we obtain

$$\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}$$

By

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k.$$

we obtain $k\lg\frac{n}{k} \leq \lg\binom{n}{k} \leq k\lg\frac{ne}{k} = k(\lg\frac{n}{k}+\lg e)$. The, as n/k grows to infinity for large n, the result follows from the two inequalities above. ■

Inequality 9.1.6
If $k = \Omega(n)$ then

$$\begin{aligned}\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k} &\Rightarrow \lg\binom{n}{k} = (1+o(1))\left(\frac{k}{n}\lg\frac{k}{n}+\frac{n-k}{n}\lg\frac{n-k}{n}\right)n\\ &= (1+o(1))nE(\frac{k}{n}),\end{aligned}$$

where $E(x) = -x\lg x-(1-x)\lg(1-x)$ is the entropy function.

Proof.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

then

$$\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}$$

Taking logarithms

$$\lg\binom{n}{k} = (1+o(1))\sqrt{\frac{n}{2\pi k(n-k}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}$$

■

Inequality 9.1.7

For any integer $n > 0$ and $k \leq n$ we have the following.

$$\binom{n}{k} \leq 2^n.$$

Inequality 9.1.8

For any integer $n > 0$ we have the following.

$$\frac{2^{2n}}{2\sqrt{n}} \leq \binom{2n}{n} \leq \frac{2^{2m}}{\sqrt{2m}}.$$

Inequality 9.1.9

The following holds.

$$\binom{2n}{n} = \frac{(2n)!}{n!\cdot n!} \geq 2^n.$$

Proof.

$$\begin{aligned}
\frac{(2n)!}{n!\cdot n!} &= \frac{(2n)\cdot(2n-1)\cdot\ldots\cdot(n+1)}{n\cdot(n-1)\cdot\ldots\cdot 1} \\
&= \frac{(2n)\cdot(2n-1)\cdot\ldots\cdot(2n-i)\cdot\ldots\cdot(2n-(n-1))}{n\cdot(n-1)\cdot\ldots\cdot(n-i)\cdot\ldots\cdot(n-(n-1))} \\
&= \frac{(2n)}{n}\cdot\frac{(2n-1)}{n-1}\cdot\ldots\cdot\frac{(2n-i)}{n-i}\cdot\ldots\cdot\frac{2n-(n-1)}{n-(n-1)} \\
&\geq 2\cdot\frac{n}{n}\cdot 2\cdot\frac{(n-1/2)}{n-1}\cdot\ldots\cdot 2\cdot\frac{(n-i/2)}{n-i}\cdot\ldots\cdot 2\cdot\frac{n-(n-1)/2}{n-(n-1)} \\
&\geq 2^n
\end{aligned}$$

The latter is due to the fact that

$$\frac{n-i/2}{n-i} \geq 1 \Leftrightarrow n-i/2 \geq n-i \Leftrightarrow i/2 \leq i.$$

■

Inequality 9.1.10 **[3]**

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.

$$\left(\frac{b-y}{a-y}\right)^y \leq \binom{a-y}{b-y}\binom{a}{b}^{-1} \leq \left(\frac{b}{a}\right)^y \leq e^{-(1-b/a)y}.$$

9.2 Series

The primary source for the listed inequalities is [1], [17].

Inequality 9.2.1 **Mercator-Newton**

For every $-1 < x \leq 1$ we have

$$\ln(1+x) = \sum_{i=1}^{\infty}(-1)^{i+1}\frac{x^i}{i} = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

Proof. Obtained from Inequality 9.2.4 after a Taylor series expansion of $\ln(x)$ for $x = 1$, and then substitution $x+1$ for x. ■

Inequality 9.2.2

For $-1 \leq x < 1$ we have the following equality by substituting $-x$ for x in Inequality (9.2.1).

$$\ln(1-x) = \sum_{i=1}^{\infty}(-1)\frac{x^i}{i} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

Moreover we obtain the following.

$$\ln(n+1) - \ln(n-1) = 2\left(\frac{1}{n} + \frac{1}{3n^3} + \frac{1}{5n^5} + \dots\right).$$

Furthermore we can also conclude that for a finite t we have the following.

$$\ln(1-x) \leq \sum_{i=1}^{t}(-1)\frac{x^i}{i},$$

and the following trivially holds.

$$\ln(1-x) \geq \sum_{i=1}^{t}(-1)\frac{x^i}{i} - \frac{x^t}{t}.$$

Inequality 9.2.3

For $x > 1/2$ we have as follows.

$$\ln(x) = \sum_{i=1}^{\infty}\frac{1}{i}\cdot\left(\frac{x-1}{x}\right)^i = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \frac{1}{3}\left(\frac{x-1}{x}\right)^3 + \dots$$

Inequality 9.2.4

For every $2 \geq x > 0$ we have the following.

$$\ln(x) \geq \sum_{i=1}^{\infty} (-1)^{i+1} (x-1)^i / i$$

Inequality 9.2.5

For we have as follows.

$$\exp(-\frac{1}{n}) = \sum_{i=1}^{\infty} (-1)^i \frac{1}{i! \cdot n^i} = 1 - \frac{1}{n} + \frac{1}{2! \cdot n^2} - \frac{1}{3! \cdot n^3} \geq 1 - \frac{1}{n}.$$

9.3 Exponential inequalities

The primary source for the listed inequalities is [1], [17].

Inequality 9.3.1

Note that for positive integer $n > 0$, we have the following.

$$(1+1/n)^n \leq e, \quad (1+1/n)^{n-1} \geq e, \quad (1-1/n)^{n-1} \geq e^{-1}.$$

They can proved by taking logarithms of both sides, then set $t = 1/n$, $t > 0$, and finally clear determine the monotonicity of a function $f(t)$ of t by calculating its first derivative and $f(0)$.

Inequality 9.3.2

The following are true.

$$e^x \geq 1 + x \quad \forall x \in R.$$

For integer $k \geq 1$, we have that

$$e^{1/k} \geq \frac{k+1}{k}, \quad \text{and} \quad e^{-1/(k+1)} \geq \frac{k}{k+1}.$$

In other words,

$$e \geq (1+1/k)^k, \quad \text{and} \quad e^{-1/(k+1)} \geq \frac{k}{k+1}.$$

Inequality 9.3.3

For all $x \in R$ and different from zero and $x < 1$.

$$\exp(-\frac{x}{1-x}) < 1 - x < \exp(-x), \quad \frac{1}{1-x} > \exp(x),$$

Furthermore,

$$x < e^x - 1 < \frac{x}{1-x}$$

Inequality 9.3.4
For every $x > -1$ we have the following.

$$\frac{x}{1+x} < (1-e^{-x}) < x.$$

For every $x > -1$ we have the following.

$$1+x > \exp(\frac{x}{1+x}).$$

For every $0 < x \leq 1.5936$ we have the following.

$$e^{-x} < 1-(x/2).$$

Inequality 9.3.5
For every $x > 0, y > 0$ we have the following.

$$e^x > (1+\frac{x}{y})^y > e^{\frac{xy}{x+y}}.$$

Inequality 9.3.6
For all $x \in R$ and different from zero and $x > -1$

$$1+x > \exp(\frac{x}{1+x}),$$

and

$$\frac{x}{1+x} < 1-\exp(-x) < x.$$

Inequality 9.3.7
It is $(1-1/N)^n \geq (1-n/N)$ for any integer $n \geq 1$ and $N > 0$.

Proof.
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

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

■

9.4 Logarithmic and other inequalities

The primary source for the listed inequalities is [1], [17], [2] , [11].

Inequality 9.4.1

For every $x > -1$ and $x \neq 0$ we have the following.

$$\frac{x}{1+x} < \ln(1+x) < x.$$

For every $x > 0$ we have the following.

$$\ln(x) \leq x - 1.$$

For every $x < 1$ and $x \neq 0$ we have the following.

$$x < -\ln(1-x) < \frac{x}{1-x}.$$

The following is true.

Inequality 9.4.2
For every $x > -1$ and $x \neq 0$ we have the following.

$$\frac{x}{1+x} < \ln(1+x) \leq \frac{x(6+x)}{6+4x} < x.$$

Inequality 9.4.3
For every $x > 0$ we have the following.

$$\ln(1+x) \geq \frac{x}{1+x/2}.$$

Proof. Consider $f(x) = \ln(1+x) - x/(1+x/2)$. It is $f(0) = 0$. Take the first derivative, and it is positive, for all $x > -1$. $f'(x) = x^2/((2+x)^2(x+1))$, therefore $f(x)$ is increasing for $x \geq 0$. ■

Inequality 9.4.4
For every $\infty > x \geq 0$ we have the following.

$$\frac{2x}{2+x} \leq \ln(1+x) \leq \frac{x(x+2)}{2 \cdot (x+1)}.$$

For every $-1 < x \leq 0$ we have the following.

$$\frac{2x}{2+x} \geq \ln(1+x) \geq \frac{x(x+2)}{2 \cdot (x+1)}$$

Inequality 9.4.5
For every $x \geq 0$ we have, as a consequence of Inequality 9.2.1, the following hold.

$$\ln(1+x) \leq x - x^2/2 + x^3/3,$$

and

$$\ln(1+x) \geq x - x^2/2.$$

Inequality 9.4.6
For every $0 < x < 0.5828$ we have the following.

$$|\ln(1-x)| < 1.5x.$$

Inequality 9.4.7
For every $0 \leq \delta < 1$ we have the following.

$$(1-\delta)\ln(1-\delta) \geq -\delta + \delta^2/2.$$

Proof. For $0 \leq x < 1$,

$$\begin{aligned} \ln(1-x) &= -\sum_{i=1}^{\infty} x^i/i \\ &= -x - x^2/2 - x^3/3 - x^4/4 - \ldots \\ (1-x)\ln(1-x) &= -x - x^2/2 - x^3/3 - x^4/4 - \ldots \\ &+ +x^2 + x^3/2 + x^4/3 + x^5/4 + \ldots \\ (1-x)\ln(1-x) &= -x + x^2/2 + x^2/6 + \ldots \end{aligned}$$

The missing terms in the last form of the equation are positive. We can then conclude

$$(1-x)\ln(1-x) \geq -x + x^2/2 + x^2/6 \geq -x + x^2/2$$

■

Inequality 9.4.8
For every $0 \leq \delta \leq 1$ we have the following.

$$(1+\delta)\ln(1+\delta) \geq \delta + \delta^2/3.$$

Proof. For $0 \leq x \leq 1$,

$$\begin{aligned} \ln(1+x) &= x - x^2/2 + x^3/3 - x^4/4 + \ldots \\ (1+x)\ln(1+x) &= x - x^2/2 + x^3/3 - x^4/4 + \ldots \\ &+ x^2 - x^3/2 + x^4/3 - x^5/4 + \ldots \\ (1+x)\ln(1+x) &\geq x + x^2/2 - x^3/6 \end{aligned}$$

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. ■

Inequality 9.4.9
For every $2e - 1 < \delta$ we have the following.

$$(1+\delta)\ln(1+\delta) \geq (2e)^{2e}.$$

Inequality 9.4.10

(a) For integer $n > 0$ the following holds.

$$\frac{1}{n+1} \leq \ln(1+\frac{1}{n}) \leq \frac{1}{n}$$

(b) For integer $n > 1$ the following holds.

$$\frac{1}{n} \leq -\ln(1-\frac{1}{n}) \leq \frac{1}{n-1}$$

Proof.
Both are also a consequence of Inequality 9.4.1. (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 $\lim_{n\to\infty}(1-\frac{1}{n})^n = 1/e$. ■

Inequality 9.4.11 **[3]**

For every $0 < t < 0.45$ we have the following.

$$\ln(1+t) > t - t^2/2 + t^3/4.$$

For every $0 < t < 0.69$ we have the following.

$$\ln(1-t) > -t - t^2.$$

For every $0 < t < 0.431$ we have the following.

$$\ln(1-t) > -t - t^2 - t^3/3.$$

9.5 Combinatorial equalities

All numbers are non-negative integers unless otherwise stated.

Lemma 9.1

For any n,k we have the following.

$$\binom{n}{k} = \binom{n}{n-k}$$

Proof. 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. ■

Lemma 9.2

For any n,k we have the following.

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

Proof. 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 $\binom{n-1}{k}$. ■

Theorem 9.1

For any $x,y \in R$, and n we have the following.

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^n y^{n-k}.$$

Corollary 9.1

We have the following.

$$(1+1)^n = 2^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Corollary 9.2

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k}(-1)^k.$$

Corollary 9.3

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.

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k}(-1)^k.$$

Proof.

$$(1-1)^n = 0 = \sum_{k=0}^{n} \binom{n}{k}(-1)^k.$$

implies

$$\binom{n}{0} + \binom{n}{2} + \ldots = \binom{n}{1} + \binom{n}{3} + \ldots$$

■

Corollary 9.4

For n and a, we have the following.

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

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

Lemma 9.3

For any n,k,m, with $k \leq r \leq n$, we have the following.

$$\binom{n}{r}\binom{r}{k} = \binom{n}{k} + \binom{n-k}{r-k}$$

Proof. 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. ■

Lemma 9.4

For any p, q, and r we have the following.

$$\sum_{k=0}^{r} \binom{a}{k}\binom{b}{r-k} = \binom{a+b}{r}$$

Proof. 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? ■

Lemma 9.5

For n and m, where $m \leq n$ we have the following.

$$\sum_{k=0}^{m} \binom{n}{k}(-1)^k = (-1)^m \binom{n-1}{m}.$$

Proof. Use induction on m. ■

Part IV

Probability problems

Chapter 10

Introductory problems

10.1 Entropy

Exercise 1.

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 by $2^y = e$ i.e. $y = \lg e$. 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 $\lg e \approx 1.442$. 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 2.

Find the minimum of $F(x) = x\lg(x) + (n-x)\lg(n-x)$, $0 < x < n$, for $n > 0$.

Solution.
For $0 < x < n$ we use $y = x/n$ that is $x = yn$, 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 3.

Analyze then entropy function $f(x) = -x\lg x - (1-x)\lg(1-x)$, with $0 < x < 1$,

Solution.

$$f(x) = -x\lg x - (1-x)\lg(1-x)$$

We have already determined that in $0 < x < 1$, function $f(x)$ has a maximum at $x = 1/2$. Then $f(x) = f(1-x)$ i.e. function $f(.)$ is symmetric around $x = 1/2$.
What is the limit $\lim_{x\to 0} f(x)$ **?**

$$\lim_{x\to 0} f(x) = \lim_{x\to 0}(-x\lg x - (1-x)\lg(1-x)) = \lim_{x\to 0}(-x\lg x - (1-0)\lg(1-0)) = \lim_{x\to 0} -x\lg x.$$

Let us change variables $y = 1/x$. Note $x \to 0$, or $1/y \to 0$, we have $y \to \infty$. Therefore

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} -x\lg x = \lim_{y\to\infty} -(1/y)\lg(1/y) = \lim_{y\to\infty} \frac{\lg y}{y} \overset{L'Hopital}{=} \lim_{y\to\infty} \frac{(\lg y)'}{(y)'} = \lim_{y\to\infty} \frac{1/y}{1} = 0^+$$

Thus the limit $\lim_{x\to 0^+} f(x) = 0$ from the right side It also suggests that $\lim_{x\to 1} f(x) = 0$. ■

10.2 Spaces

Exercise 4.

An (S,T) pair 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 5.
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 $a=0,b=1$. Consider also $\cap_{m\in N}(a-\frac{1}{m},a+\frac{1}{m})=[a,a]=\{a\}$.

■

Exercise 6.
(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

$$S = S \cup \emptyset \Rightarrow P(S) = P(S) + P(\emptyset)$$

Given 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. ■

Chapter 11

Probability of events

11.1 Events and their properties

Exercise 7.

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.

Show that 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.

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 8.
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 9.

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 10.
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 11.

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.

$$\begin{aligned}
|\cup_{i=1}^{n+1} A_i| &= |(\cup_{i=1}^{n} A_i) \cup A_{n+1}| \\
&= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |(\cup_{i=1}^{n} A_i) \cap A_{n+1}| \\
&= |\cup_{i=1}^{n} A_i| + |A_{n+1}| - |\cup_{i=1}^{n} (A_i \cap A_{n+1})| \\
&= \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|\cap_{i\in I} A_i| + |A_{n+1}| - \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|(\cap_{i\in I} A_i \cap A_{n+1})| \\
&= \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|\cap_{i\in I} A_i| + |A_{n+1}| - \sum_{I\subseteq\{1,\ldots,n\}} (-1)^{|I|+1}|(\cap_{i\in I\cup\{n+1\}} A_i)| \\
&= \sum_{I\subseteq\{1,\ldots,n+1\}, n+1\notin I} (-1)^{|I|+1}|\cap_{i\in I} A_i| + \sum_{I\subseteq\{1,\ldots,n+1\}, n+1\in I} (-1)^{|I|+1}|(\cap_{i\in I} A_i)| \\
&= \sum_{I\subseteq\{1,\ldots,n+1\}} (-1)^{|I|+1}|\cap_{i\in I} A_i|
\end{aligned}$$

■

11.2 Toys: dies, coins and cards

11.2.1 Throw or roll a die or dice

Exercise 12.
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 13.
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 14.
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 15.
(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) $6!/6^6$. ■

11.2.2 Toss a coin

Exercise 16.

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 17.
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 18.

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 19.
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 20.
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 21.

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 22.
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 23.
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 24.
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 25.
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

$$E(N)=1\cdot(1/36)+2\cdot(15/36)+3\cdot(5/9)=91/36.$$

■

11.2.3 Deck of cards

Exercise 26.

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 27.
(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

$$\frac{4\cdot 3\cdot 2\cdot \left(48!/(12!)^4\right)}{52!/(13!)^4}.$$

■

11.3 Conditional probabilities

Exercise 28.

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 29.

(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 30.

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 31.
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 32.
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. ■

11.4 Independence of events

Exercise 33.

(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.

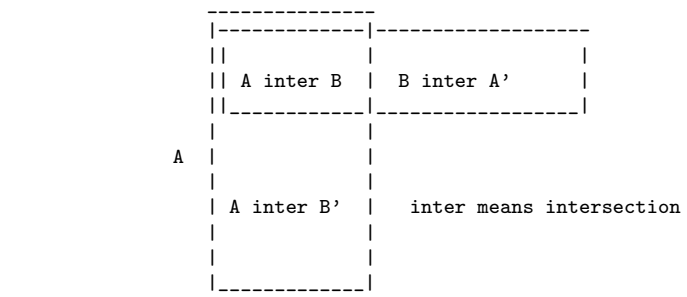
Solution.
Note that $A = (A \cap B) \cup (A \cap B')$. Then

$$P(A) = P(A \cap B) + P(A \cap B') = P(A)P(B) + P(A \cap B').$$

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 34.

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.

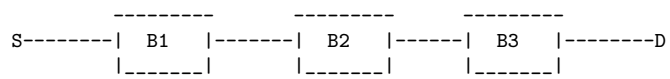
(b) We calculate $P(A\cap B\cap C)=1/4$. But $P(A)P(B)P(C)=(1/2)(1/2)(1/2)=1/8$. Therefore

$$P(A\cap B\cap C)\neq P(A)P(B)P(C).$$

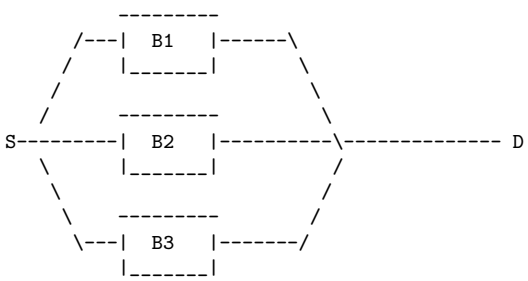
■

Exercise 35.

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 36.
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 37.
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 38.
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 39.

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 40.

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 41.
Two statistics 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 42.
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 43.
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 44.
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 45.
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 46.

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 47.
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

$$0.75 + 0.25 * 0.75 + (0.25)^2 0.75 + (0.25)^3 0.75 = 0.75(1 + 0.25 + 0.25^2 + 0.25^3) = 0.99609375.$$

■

Chapter 12

Random variables

12.1 Discrete random variables

Exercise 48.
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 49.
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 3.5 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 $a,b \in \mathbb{R}$.

$$\begin{aligned}
E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax+by) f_{X+Y}(x,y)\,dxdy \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X+Y}(x,y)\,dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X+Y}(x,y)\,dxdy \\
&= a\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X+Y}(x,y)\,dy\right)dx + b\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X+Y}(x,y)\,dx\right)dy \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} y f_Y(y)\,dy \\
&= aE(X)+bE(Y).
\end{aligned}$$

■

Exercise 50.
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 51.
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 3.5.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} y f_Y(y)\int_{-\infty}^{\infty} x f_X(x)dx \\ &= E(X)E(Y). \end{aligned}$$

If X,Y are discrete and X,Y independent, $X \neq 0$, we have the following using Theorem 3.5 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 52.

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 53.
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.

$$a \leq X \leq b \Rightarrow 0 \leq X-a \leq b-a \Rightarrow 0 \leq X-a \leq b-a \Rightarrow 0 \leq \frac{X-a}{b-a} \leq 1 \Rightarrow 0 \leq Y \leq 1.$$

■

12.2 Continuous random variables

Exercise 54.

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 55.
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 56.
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 57.
(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 58.
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 59.

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 60.
(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 61.
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 62.
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 $X \sim N(0,1)$. See a later exercise and its solution. ■

Exercise 63.

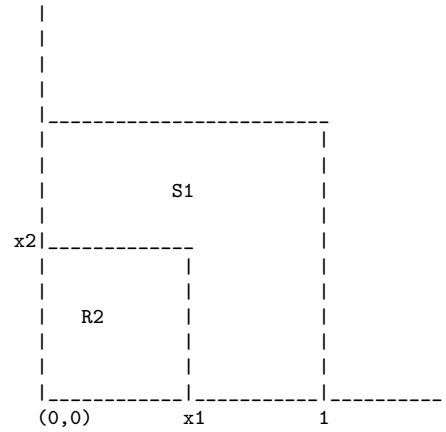
Let (x_1, x_2) be a point chosen uniformly at random in the unit square with coordinates $(0,0),(1,0),(1,1),(0,1)$ counterclock-wise. Form r.v. $X = x_1^2$, $Y = x_2^2$ and $Z = x_1 + x_2$. Show that X, Y are independent but Y, Z are not by determining the $F_{X,Y}(x,y)$ and $F_{Y,Z}(x,z)$.

Solution.
We note the following.

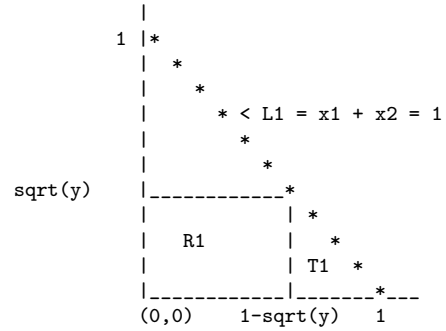
$$F_X(x) = P(X \leq x) = P(x_1^2 \leq x) = P(x_1 \leq \sqrt{x}) = F_{X_1}(\sqrt{x}) = \sqrt{x}.$$

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.



■

Exercise 64.

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. ■

12.3 Birthdays

Exercise 65.

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 66.

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 67.

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 68.
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. ■

12.4 Miscellanea

Exercise 69.
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 70.

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 71.

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 72.

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 73.

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 74.

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 75.

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

$$\begin{aligned} E(X) &= \sum_{i=1} nP(X=i) \cdot i \\ &= \sum_{i=1} n \frac{2^{i-1}}{2^n-1} \cdot i \\ &= \frac{1}{2^n-1} \cdot \sum_{i=1} n 2^{i-1} \cdot i \\ &= \frac{1}{2^n-1}(n2^n - 2^n + 1) \\ &= (n-1) + \frac{n}{2^n-1} = (n-1) + \frac{n}{N}. \end{aligned}$$

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 76.
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 77.

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 $\sum d_i a_i \equiv 0 \mod p$? 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. $\sum d_i a_i \equiv 0 \mod p$.

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \mod p \\ d_1 a_1 + d_2 a_2 + \ldots + d_n a_n &\equiv 0 \mod p \\ d_1 a_1 &\equiv (-d_2 a_2 - \ldots - d_n a_n) \mod p \\ d_1 a_1 &\equiv Z \mod p \end{aligned}$$

where $Z \equiv (-d_2 a_2 - \ldots - d_n a_n) \mod p$. Then,

$$\begin{aligned} \sum_i d_i a_i &\equiv 0 \mod p \\ d_1 a_1 &\equiv Z \mod p \\ a_1 &\equiv (d_1)^{-1} Z \mod p \\ a_1 &\equiv B \mod p \end{aligned}$$

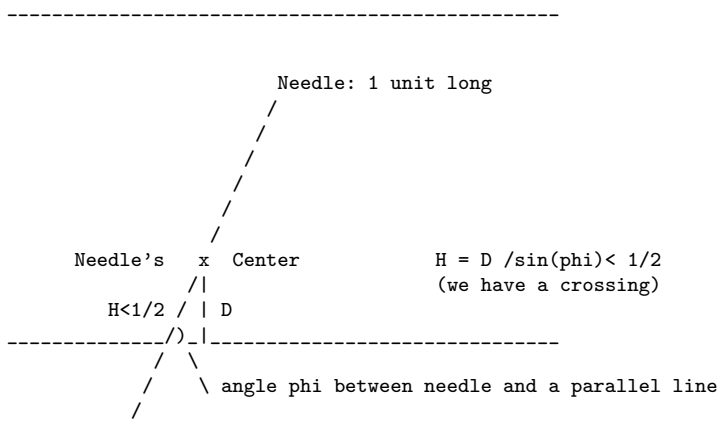
The inverse of d_1 exists since $d_1 x \equiv 1 \mod p$ has a single solution for x by the fact that $d_1 < p$ is such that $(d_1, p) = 1$ and p is prime. $B = (d_1)^{-1} Z \mod p$ is an integer in $0, \ldots, p-1$. a_1 is a (uniformly at) random (chosen) 0,1. What is the probability that the random 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 78.

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 79.

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. ■

12.5 Permutations and Derangements

Exercise 80.
(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 81.

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 82.

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 83.
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 Y denote the number of fixed points.

$$Y = X_1 + X_2 + \ldots + X_n \Rightarrow E(Y) = E(X_1 + X_2 + \ldots + X_n) = \sum_E (X_i) = nE(X_i) = n(1/n) = 1.$$

We used $E(X_i) = 1/n$. ■

Exercise 84.
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 85.
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.

$$E(R)=\int_0^1 rf_R(r)dr=\int_0^1 r2rdr=\int_0^1 d2r^3/3==2/3$$

■

12.6 Bin and Balls

Exercise 86.
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 87.
(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 88.
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 $\left\{ \begin{array}{c} N \\ n \end{array} \right\}$.
(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! ■

12.6.1 ... or not

Exercise 89.

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 90.

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 91.

Assume $n = N$. What is the probability that bin k has at least t balls in it?
We provide the following definition

Definition 12.1

We say an event E dependent on n occurs with high probability, if $P(E) \leq 1 - 1/n^c$ for some constant $c > 1$.

Instead of writing "with high probability" we shall write "w.h.p." or "whp" instead.

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 (12.2) and (12.3) appears in Feller [9] and in its complete form including Equation (12.1) as Theorem 1.1 in Bóllobás [3], where $B(n,p;m)$ is expanded and bounded,

Theorem 12.1 **Feller[9], Bóllobás[3]**

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 $m = \lceil upn \rceil$, $u > 1$,

$$P(S_n \geq m) \leq \frac{u}{u-1} B(n,p;m). \tag{12.1}$$

Furthermore,

$$P(S_n \geq m) \leq \frac{m(1-p)}{(m-np)^2}. \tag{12.2}$$

and

$$P(S_n \leq m) \leq \frac{(n-m)p}{(np-m)^2}. \tag{12.3}$$

The first form appears in [3]; the other two in [9].

By Eq.(12.1) with $t = \lceil upn \rceil$, $p = 1/n$ and $t = \lceil u \rceil$, $u > 1$. In Eq.(12.1) the $u/(u-1)$ becomes at most 1.5 for $u > 3$ or equivalently $t > 3$. Therefore for $t > 3$

$$P(Bin\ k \geq t) = P(S_n \geq t) \leq \frac{u}{u-1} B(n,p;t) \leq 1.5 \binom{n}{t} (1/n)^t. \tag{12.4}$$

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).

$$P(Bin\ k \geq t) \leq \sum_{S_n \subseteq \{1,\ldots,n\}, |S_n| = t} \prod_i P(\text{ball i } \in S_n \text{is in bin } k) \leq \binom{n}{t} p^t = \binom{n}{t} (1/n)^t \tag{12.5}$$

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.(12.4) and Eq.(12.5) is a 1.5. We proceed discarding 1.5 as follows.

Using inequality 9.1.3 we obtain from Eq.(12.5) the following.

$$P(Bin\ k \geq t) \leq \left(\frac{ne}{tn}\right)^t \leq \left(\frac{e}{t}\right)^t.$$

Assume $n \geq e^2$ 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 92.

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 93.

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 94.
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 95.

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 96.

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.

$$P(\max_i X_i \geq k) = P(\exists i : X_i \geq k) \leq \sum_i P(X_i \geq k) = nP(X_i \geq k).$$

■

12.7 Collecting coupons

Exercise 97.

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 98.

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

$$E(X) = \sum_k E(X_k) = \sum_k \frac{n}{n-k+1} = n\sum_k \frac{1}{k} \approx n \ln n.$$

■

Chapter 13

Distributions

13.1 Distributions

13.1.1 Binomial process and Bernoulli trials

Exercise 99.
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 100.

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 101.
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 102.
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 103.

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 $(n+1)p - 1 < m \leq (n+1)p$. ■

Exercise 104.
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 $a/(1-a) = rq/(np-r)$. ■

Exercise 105.

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 106.

We would like to determine whether $k \in A$ or not. We would attempt to examine the problem from a deterministic and randomized point of view.
(a) Give a deterministic algorithm for this problem and its solves this problem deterministically in time $O(n)$ is linear search.
(b) Say there is a randomized algorithm that determines membership in A i.e. whether $k \in A$ or $k \notin A$. The randomized algorithm performs a test $T(k)$ that take time $\Theta(1)$ i.e. constant time. If $k \in A$ then $T(k)$ is always a YES indicating indeed $k \in A$. If $k \notin A$ then $T(k)$ is a YES appears with probability $1-p$ indicating $k \in A$ (false positive) and a NO appears with probability p indicating $k \notin A$ correctly. A test $T(k)$ is independent of any previous tests involving k or other keys. Thus it is possible for the same key k to have a YES followed by a NO. The randomized algorithm is repeated m times for each key k. What is its running time?
(c) if $k \notin A$ what is the probability that k passes all m tests?
(d) Let Q be some fixed parameter. How large should m be so that a key k is incorrectly labeled is at most Q? (You may assume $0 < Q < 1$.)
(e) How many tests do we need until we can conclude with cetainty $k \notin A$? Give the expected number of tests $E[T]$.

Solution.
(b) Obviously $\Theta(m)$.
(c) It is $(1-p)^m$.
(d) The probability for a key k such that $k \notin A$ we have to have a false positive conclusion is $(1-p)^m$. THis is $(1-p)^m \leq Q$ for $m \leq \lg Q / \lg(1-p)$. Thus for $m > \lg Q / \lg(1-p)$ the probability of false positive is Q or less.
(e) This is a geometric distribution. The probability that we have a proof of non-membership in i tests is $P(T=i) = (1-p)^{i-1}p$. Then the expected number of tests $E[T]$ is

$$\begin{aligned} E[T] &= \sum_{i=0}^{\infty} i(1-p)^{i-1}p \\ &= p/(1-p) \cdot \sum_{i=0}^{\infty} i(1-p)^i \\ &= \frac{p}{1-p} \cdot \frac{1-p}{(1-(1-p))^2} \\ &= \frac{p}{1-p} \cdot \frac{1-p}{p^2} \\ &= \frac{1}{p}. \end{aligned}$$

The latter step come from another problem when $x \to \infty$ and $x < 1$. We then substitute $1-p$ for x. ■

Exercise 107.

The following question relates to the results of the previous problem.
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 $1/p_T + 1/p_H = 4$. ■

13.2 Randomness

Exercise 108.

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 109.

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 110.

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.

■

Exercise 111.
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 112.

```
   PermutE(A,n) // A is an array A[1..n]
1.   for(i=1;i<=n;i++) {
2.       //random(1,n)  returns
3.       //a uniformly at random integer between 1 and n
4.     swap(A[i], A[random(1,n)]);
5.   }
6.   return(A);
```

(a) Does PermutE(A,n) generate a permutation?
(b) Does PermutE(A,n) generate a random permutation?

Solution.
YES and NO.
(a) YES, one can show easily that a permutation is returned.
(b) NO. Algorithm `PermutE` generates permutations but not uniformly at random. Consider for example the case of $n = 3$. On 3 keys, there are $3! = 6$ different permutations. Algorithm `PermutE` can generate $3^3 = 27$ possible outcomes each one corresponding to a permutation; however not all of them are distinct from each other. Given that there are only 6 distinct permutations on 3 items, and that 6 does not divide 27, some of these 6 permutations will be generated more often than the others, i.e. there is a bias towards certain permutations over others. If you exhaustively generate all possible 27 outcomes of the algorithm, you can see that three permutations show up 5 times and 3 show up 4 times, i.e. there is a bias in favor of certain permutations such as $(2,3,1)$. You can see in the drawing below the possible outcomes after only the first two swaps have been performed. Out of the 9 possible permutations that can be generated so far $(2,1,3)$, $(1,2,3)$ and $(2,3,1)$ appear twice already. In general the "decision tree" corresponding to the outcomes has n^n leaves, each one reached with the same probability. The number of permutations for n items is however $n!$. The leaves can not be divided equally among the permutations since n^n is not a multiple of $n!$ for $n > 2$. Thus each permuation is not available with the same equal probability $1/n!$ and thus the algorithm is incorrect.

```
Starting Input A[i]=i for i=1,2,3

 1                            9             27

          1st           2nd (2,1,3)
        ------> (1,2,3) --- (1,2,3)
        |                   (1,3,2)
        |
        | 1st           2nd (1,2,3)   3rd  not shown
(1,2,3) ------> (2,1,3) --- (2,1,3)  ....  and so on    ...
        |                   (2,3,1)
        |
        | 1st           2nd (2,3,1)
        ------> (3,2,1) --- (3,2,1)
                            (3,1,2)
1st = Possible outcomes after swap ( A[1], A[random(1,3)] ) is done
2nd = Possible outcomes after swap ( A[2], A[random(1,3)] ) is done
3rd = Possible outcomes after swap ( A[3], A[random(1,3)] ) is done
```

■

Exercise 113.

Consider the following algorithm.

```
  FunPermute(A,n) // A is an array A[1..n]
1. for(i=1; i<= n-1 ; i++){
2. swap(A[i], A[random(i+1,n)];
3. }
4. return(A)
```

Does this code generate any permutation ? Does it produce a uniform at random non-identity permutation?

Solution.
No or not if $n > 2$. The element in the first position is swapped out at $i = 1$; in the remainder it cannot be re-occupy $A[1]$.

■

Exercise 114.

You are given RAND(a,b) that returns a random integer number between a (inclusive) and b (inclusive) with uniform probability. Show how to use this RAND function to generate a random permutation in $O(n)$ time of n distinct objects stored in array $X[1..n]$. You may assume that a call to RAND takes constant time $O(1)$.

For example if X has three items $< 10,20,30 >$ then after your algorithm is run it is equally likely that the output is $< 10,20,30 >$ or $< 20,10,30 >$ or any one of the remaining 4 permutations of the three distinct objects 10, 20, 30.

Solution.

```
   Permute(A,n) // A is an array A[1..n]
1.   for(i=1;i<=n;i++) {
2.       //random(1,n)  returns
3.       //a uniformly at random integer between 1 and n
4.     swap(A[i], A[random(i,n)]);
5.   }
6.   return(A);
  or
   Permute(A,n) // A is an array A[1..n]
1.   for(i=n; i>=1 ;i--) {
4.     swap(A[i], A[random(1,i)]);
5.   }
6.   return(A);
```

The algorithm (second version) works in-place. At the start, at the conclusion of iteration $I = 1$, where $i = n+1-I$, the items at index $i = n+1-I$ stores any one of the items in position $n+1-I, n-I, \ldots, 1$, a choice among $n+1-I = n$ items. By induction, in iteration I, where $i = n+1-I$, the items at index $i = n+1-I$ stores any one of the items in positions $n+1-I, n-I, \ldots 1$, a choice among i of them. Therefore for iteration $I = 1$ (equivalent to $i = n$), there are n outcomes for $A[i]$, for iteration $I = 2$ (equivalent to $i = n-1$) there are $n-1$ outcomes for $A[i]$), and so on. The decision tree has $n!$ leaves, corresponding to the outcomes of Permute. Each of the outcomes is generate with the same probability $1/(n(n-1)\ldots 1) = 1/n!$.
The running time is obviously $O(n)$ if a single call to RAND is $O(1)$.
Inductive hypothesis: Before the start of iteration i, $A[1..i-1]$ contains an $(i-1)$-permutation ($i-1$ of the n keys) with probability $(n-i+1)!/n!$. (For a base case this is true as well by default.)
Inductive step. An i-permuations is an $(i-1)$-permutation followed, in the i-th iteration by the element picked for $A[i]$. Let the $(i-1)$-permuation of the induction step be p. Its probability of occurence is by the inductive hypothesis $(n-i+1)!/n!$. The even that a given item is picked among the $A[i..n]$ and swapped at $A[i]$ is $1/(n-i+1)$. Thus the probability that the i-permuation generated at the end of the i iteration is as is is $1/(n-i+1) \cdot (n-i+1)!/n! = (n-i)!/n!$. At the completion of the $i = n$ iteration, a given permutation formed appears with probability $(n-n)!/n! = 1/n!$, as required. ■

Exercise 115.

The `Higlander Casino` plays a game of random permutations on 4 items as follows.

- The client bets on a permutation of his/her choice.
- The casino uses the incorrect one of the algorithms (i.e. PermutE) to generate a random permutation.
- The casino pays the client the fair amount of money if the client won (i.e. guessed right). Fair money means in the long run neither the client nor the casino would make any money IF their algorithm chose correctly and fairly a permutation (i.e. for n items the casino pays n dollars for every dollar bet right).

Since the casino's algorithm is incorrect, what methodology could you follow to gain over the casino? What edge (percentage-wise) do you expect to gain over the casino? Explain.

Solution.
The casino is using PermutE. With 4 items the decision tree has $4^4 = 256$ leaves. There are however $4! = 24$ permutations on 4 items and these 24 permutations correspond to the 256 leaves. So, on the average $256/24$ leaves correspond to each permutation (or a given permutation appears on the average that often as outcome of PermutE). The ratio $256/24$ is equal to 10.666. A permutation can appear as a leaf 10 or 11 times but not 10.666 times. Therefore there must be a permutation that appears at least 11 times since otherwise all permutations appear only 10 times and therefore the 24 of them can only appear $24 \times 10 = 240$ times as leaves but there are 256 leaves permutation generated by PermutE. Therefore our algorithm in beating the casino is to identify this permutation (keeping track of what shows up as a result of successive games and histograming these outcomes) and after doing so, to bet on it time after time. Our expected probability of winning would thus be (at least) $11/256$. Since the casino pays fair money, if we bet 1 we win 24\$. So our expected return would be $11/256 * 24 = 264/256 = 1.03125$. So our edge over the casino is 3.125%.

■

Exercise 116.

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 (3 points)?
(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 (17 points).
(c) Let d be a vector of integers mod p (i.e. items 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 117.

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. ■

Exercise 118.

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.

■

13.3 Ramsey numbers

Exercise 119.

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.

a) 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 120.

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. (Replace 3 by e noting that the base case is still true.)

■

Exercise 121.

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 2.1 the probability of a union of events is at most their sum of probabilities. Thus

$$P(\text{Graph has k-clique or k independent set } \leq 2\binom{n}{k}2^{-\binom{k}{2}}.$$

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 122.

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.

$$\begin{aligned}
n^k(1-\frac{1}{n})^{\frac{k^2}{4}} &\leq \frac{5}{6} \Rightarrow \\
n^k(1-\frac{1}{n})^{\frac{k^2 n}{4n}} &\leq \frac{5}{6} \Rightarrow \\
n^k e^{-\frac{k^2}{4n}} &\leq \frac{5}{6} \Rightarrow \\
k\ln n + \ln\frac{6}{5} &\leq \frac{k^2}{4n} \Rightarrow \\
\ln n &\leq \frac{k}{n} \Rightarrow \\
\ln c + \ln k - \ln\ln k &\leq \frac{\ln k}{c}
\end{aligned}$$

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

$$\ln n \leq kn^{-\frac{2}{3}} \Rightarrow k \geq n^{\frac{2}{3}}\ln n$$

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 123.

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 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,

$$\begin{aligned}
\Pr(G \text{ has a } k\text{-noisy clique}) &\leq \binom{n}{k} \Pr(\text{A particular set of } k \text{ vertices is a noisy clique}) \\
&\leq \binom{n}{k} \binom{\binom{k}{2}}{k^2/10} 2^{-\left(\binom{k}{2} - k^2/10\right)} \\
&\leq \binom{n}{k} \binom{k^2/2}{k^2/10} 2^{-\left(\frac{k(k-1)}{2} - \frac{k^2}{10}\right)} \\
&\leq n^k \binom{k^2/2}{k^2/10} 2^{-.4k^2 + k/2}
\end{aligned}$$

In order to bound this, observe the following noting that 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 larger, chosing $\varepsilon = .02$ suffices. (One could get a better bound by a more careful approximation or approximation arguments.) ■

13.4 Van der Waerden

Exercise 124.

(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.

Observe that 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 125.
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 $r \geq 2^{n-1}$.

■

Chapter 14

Generating functions

14.1 Generating functions

Exercise 126.
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 127.
(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 128.
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 129.
(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 130.
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 131.
(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 132.
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 133.
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 134.
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 135.
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 136.
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 137.
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 138.
Let X,Y follow a geometric distribution and are independent. Thus $X \sim G(p)$ and $Y \sim G(p)$. Then

$$f_X(k) = f_Y(k) = q^{k-1}p.$$

Let $Z = X + Y$. Find the p.d.f of Z.

Solution.
We have $m_{X+Y}(t) = m_X(t)m_Y(t) = \frac{p^2e^{2t}}{(1-qe^t)^2}$. Moreover $p_{X+Y}(t) = (p^2t^2)/(1-qt)^2$. Be reminded, for $|x| < 1$, we have the following

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, \quad \sum_{i=0}^{\infty} ix^i = \frac{x}{(1-x)^2}, \quad \sum_{i=0}^{\infty} ix^{i+1} = \frac{x^2}{(1-x)^2}.$$

Then substitute in the last sum $x = qt$.

$$p_{X+Y}(t) = (p^2t^2)/(1-qt)^2 = \frac{p^2}{q^2}\sum_{k=0}^{\infty} i(qt)^{i+1}.$$

The coefficient of t^k is

$$f_{X+Y}(k) = \frac{p^2}{q^2}(k-1)q^k.$$

■

Exercise 139.
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 140.
Find the characteristic function of $X \sim Bernoulli(p)$.

Solution.

$$\begin{aligned} c_X(t) &= E(e^{itX}) \\ &= e^{it\cdot 1}\cdot p + e^{it\cdot 0}\cdot(1-p) \\ &= pe^{it} + (1-p) \\ &= p\cos t + ip\sin t + (1-p). \end{aligned}$$

■

Exercise 141.

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 142.
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,

$$F_X'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

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 $E(X) = 0$. ■

Chapter 15

Probability inequalities

15.1 Probability inequalities

Exercise 143.
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 144.

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 145.

(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 146.

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 147.

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.

$$P(S_n \geq r) \leq P(\cup E_S) \leq \sum_{S,|S|=r} P(E_S) \leq \binom{n}{r} p^r. \tag{15.1}$$

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.(15.1) 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 148.

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) We first utilize the Chernoff bound of Corollary 15.1. We pick $\delta = 2\sqrt{n} > 2e-1$ for sufficienly 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 15.1 with $r = 2/sqrtn$. The bound is an even better $\approx 1/n^{\sqrt{n}}$. Hoeffding provides the same bound through Theorem 15.1. Corollary 15.2 or Corollary 15.3 of Hoeffding's theorem provide lesser bound $\exp(-8n)$.
(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 149.

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 15.1

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 $r = (1+\delta)p$, $\delta > 0$.

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} \tag{15.2}$$

The following Corollary can be obtained from Lemma 15.1. It provides a more tangible upper bound than the generic one of the Lemma.

Corollary 15.1

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,

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq 2^{-(1+\delta)pn}. \tag{15.3}$$

We first utilize the Chernoff bound of Corollary 15.1. 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 15.1 with $r = 2/\sqrt{n}$. The bound is an even better $\approx 1/n^{\sqrt{n}}$. Hoeffding provides the same bound through Theorem 15.1. Corollary 15.2 or Corollary 15.3 of Hoeffding's theorem provide a lesser bound of $\exp(-8n)$.

Theorem 15.1

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.

$$\begin{aligned} P(\bar{X} - p \geq h)) &= P(S_n \geq (p+h)n) \\ &\leq \exp(-D((p+h)||p)n) \\ &\leq \left[\left(\frac{p}{p+h} \right)^{p+h} \left(\frac{1-p}{1-p-h} \right)^{1-p-h} \right]^n . \end{aligned} \tag{15.4}$$

Corollary 15.2

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.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-2n^2\delta^2 p^2}{n(b-a)^2} \right). \tag{15.5}$$

Corollary 15.3 appears in [15] as 5.3 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 15.3

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.

$$P(S_n \geq pn + h) \leq \exp\left(\frac{-2h^2}{\sum_{i=1}^n (b_i - a_i)^2} \right). \tag{15.6}$$

(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 150.

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 15.4.

Corollary 15.4

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,

$$P(|S_n - np| \geq \delta pn) \leq 2 \cdot \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{15.7}$$

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.

Chapter 16

Miscellaneous problems

16.1 Hashing

Exercise 151.
(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 152.

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 16.1 **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 153.

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 equation follows. 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 154.

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 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

$$\begin{aligned} \ln Z &= \ln(e^k/k^k) = k - k\ln k \\ &= \frac{A\ln n}{\ln\ln n} - \frac{A\ln n}{\ln\ln n} \cdot \ln fracA\ln n\ln\ln n \\ &= \frac{A\ln n}{\ln\ln n} - \frac{A\ln n}{\ln\ln n} \cdot (\ln A + \ln\ln n - \ln\ln\ln n) \\ &= A(1-\ln A)\frac{\ln n}{\ln\ln n} - A\ln n + A\ln n \cdot \frac{\ln\ln\ln n}{\ln\ln n} \\ &= A(1-\ln A)\frac{\ln n}{\ln\ln n} - (A-1)\ln n + A\ln n \cdot \left(-1 + \frac{\ln\ln\ln n}{\ln\ln n}\right) \\ &= Q_1 - (A-1)\ln n + Q_2. \end{aligned}$$

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 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 $E[D] < k+1$. ■

Exercise 155.

In open addressing under the uniform hashing assumption, let $p(n,N)$ be the probability that there are no collisions, during the insertion of n keys into a hash table T with N slots that is originally empty.
Show by induction that

$$p(n,N) \leq \exp(-\frac{n(n-1}{2N}).$$

Deduce that for $n > \sqrt{N}$ the probability of having a collision grows to close to one.
Note that $\exp x = e^x \geq 1+x$ for any real x.

Solution.
Consider $p(n+1,N)$ and argue similarly to the derivation of the number of probes for Insertion. For the $n+1$-st key not to cause any collision in a hash table with n keys inserted with no collisions, it must hit an empty slot an event that occurs with probability $1-n/N$. This assumes that the insertion of the $n+1$-st key is independent of the previous insertions. Therefore $p(n+1,N) = (1-n/N)p(n,N)$, since the probability that the insertion of say $n+1$ cause not collision is the probability that the insertion of the $n+1$-st key causes no collisions and the prior insertion of the remaining n keys caused no collisions; the latter is $p(n,N)$. Therefore

$$\begin{aligned} p(n+1,N) &= (1-n/N)p(n,N) \\ &= (1-n/N)(1-(n-1)/N)p(n-1,N) \\ &= (1-n/N)(1-(n-1)/N)\ldots(1-1/N)(1-0/N)p(0,N) \\ &= (1-n/N)(1-(n-1)/m)\ldots(1-1/N) \end{aligned}$$

Since $e^x \geq 1+x$, setting $x = -i/N$ we obtain that $1-i/N \leq e^{-i/N}$. Therefore

$$\begin{aligned} p(n+1,N) &= (1-n/N)(1-(n-1)/N)\ldots(1-1/N) \\ &\leq \exp(-n/N)\exp(-(n-1)/N)\ldots\exp(-1/N) \\ &\leq \exp(-(1+2+\ldots+n)/N) \\ &= \exp(-n(n+1)/2N) \end{aligned}$$

Therefore $p(n,N) \leq \exp(-(n-1)n/(2N))$ as required. Consider $n \gg \sqrt{N}$, the $n(n-1)/2N \gg 1$, and thus $p(n,N)$ is upper bounded by $\exp(-A)$, where A is large and positive. However $\exp(-A)$ goes to zero and thus $p(n,N)$ goes to zero as well.
Note. One might remark that the probability of having a collision is $0+1/N+2/N+\ldots+n/N = n(n-1)/(2N)$. This is not correct. This bound is not exact but an upper bound. Thus, if one uses

$$P(\text{Nocollision}) + P(\text{Collision}) = 1$$

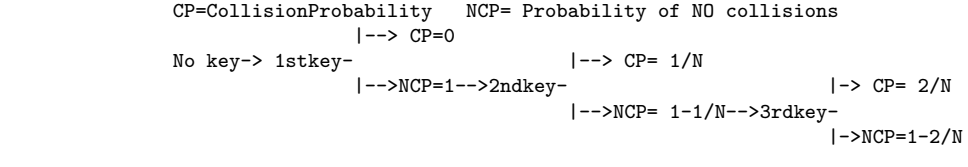
given that

$$P(\text{Collision}) \leq n(n-1)/(2N)$$

one can derive a lower bound for

$$P(\text{Nocollision}) \geq 1 - n(n-1)/(2N)$$

However this is not useful. To understand why the bound above is not exact think of the case of three keys only.



Think of the diagram as a binary tree with root on the left. Let's see for what probability we reach a leaf a CP or NCP and add the probabilities of reaching all CP leaves and all NCP leaves. There is one NCP leaf on the rightmost side. We reach this with probability $(1-1/N)(1-2/N)$ which is the probability of no collisions after three keys have been inserted. The probability bound is the product of probabilities on all intermediate nodes. There are three CP nodes. The topmost one is 0, the second one is reached with probability $1/N$ and the last one with probability $(1-1/N)2/N$. Thus the probability of having a collision after three keys are inserted is $1/N+2/N-(1/N)(2/N)$ which is close to $1/N+2/N$ but not equal to. This problem is a variant of a problem known as the Birthday problem. Let $N=365$ (forget about leap years). Suppose you have n people in a room. What is the probability of having no two with the same birthday? For what value of n is the probability of having two people with the same birthday close to 1? One can see that for $n=\sqrt{N}$ collisions start to show up with a significant (non negligible) probability, i.e. people exist with the same birthdays.
This problem is also known as the Birthday problem. Let $m=365$ (forget about leap years). Suppose you have n people in a room. What is the probability of having no two with the same birthday? For what value of n is the probability of having two people with the same birthday close to 1? One can see that for $n=\sqrt{m}$ collisions start to show up with a significant (non negligible) probability, i.e. people exist with the same birthday. ■

16.2 Hashing

Exercise 156.
(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 157.

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 16.2 **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 $E[D] = n(1-1/n)^{n-1} \approx n^2/(e(n-1))$. ■

16.3 Miscellanea

Exercise 158.

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 159.

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 160.

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 161.

Let A be an $n \times n$ matrix of zeroes and ones. Let b be an $n \times 1$ vector of -1 and $+1$ Let $y = A \times b$ be a matrix-vector product. Let vector b's entries are uniformly at random with $p = 1/2$ between -1 and $+1$

We would like to bound

$$P(\max x_i \geq \sqrt{B}) = O(1/n)$$

What is the value of B that achieves this?

Solution.

Consider $a_i = \sum_j a_{ij}$. Therefore a_i is the sum of ones of row i of A. If $a_i < \sqrt{B}$ then $\sum_j a_{ij} b_j \leq \sqrt{B}$ with probability 1 as there are not enough ones in row i of A. Otherwise $a_i \geq \sqrt{B}$. ■

16.4 Random graphs

Exercise 162.

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 163.

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 164.
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 165.

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 166.

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

$$\begin{aligned}
\text{Pr(there is no p.m.)} &\leq \sum_{i=1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

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 167.

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 $p = \frac{(1+\varepsilon)\log n}{n}$ where $\varepsilon > 0$ (the logarithms here are natural ones), 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 therefore

$$(1-p)^n \leq e^{-pn} = \frac{1}{n^{1+\varepsilon}},$$

while $1/(1-p) < 2$. We also use $\binom{n-1}{k} \leq n^k$.

$$\begin{aligned} P &= n\sum_{k=0}^{10} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &\leq n(1-p)^n \sum_{k=0}^{10} n^k \frac{(1+\varepsilon)^k \log^k n}{n^k (1-p)^{k+1}} \\ &\leq n\frac{1}{n^{1+\varepsilon}} \sum_{k=0}^{10} (1+\varepsilon)^k \log^k n \; 2^{k+1} \\ &\leq \frac{11 \cdot 2^{11} (1+\varepsilon)^{10} \log^{10} n}{n^{\varepsilon}} \to 0 \text{ when } n \to \infty. \end{aligned}$$

The sum was bounded above by 11 times its largest term. ■

Exercise 168.

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 vertex graph 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 169.

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

$$\begin{aligned}
\text{Pr(there is no p.m.)} &\leq \sum_{i=1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \sum_{i=n/2+1}^{n} \binom{n}{i}\binom{n}{n-i+1}(0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

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 170.

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 above 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 above 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 171.

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 them. 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 172.

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-vertex cycles 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 173.

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) This part has been solved earlier. Once again, 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). ■

16.5 Random matrices

Exercise 174.

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 count 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 employ Stirling's approximation formula. We get an upperbound by ignoring the square root terms.

$$\left(\begin{array}{c} n \\ \frac{n}{2} \end{array} \right)^2 (1-\frac{k}{n})^{\frac{n^2}{4}} \approx \left(\frac{(\frac{n}{e})^n}{(\frac{n}{2e})^{n/2}(\frac{n}{2e})^{n/2}} \right)^2 (1-\frac{k}{n})^{\frac{n}{k}\frac{kn}{4}} \leq 4^n e^{-\frac{kn}{4}} \tag{16.1}$$

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 the equation 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 175.

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 $\frac{49n}{100}$ 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

$$\left(\begin{array}{c} n \\ 2 \end{array} \right) \sum_{i=0}^{\frac{49n}{100}} \left(\begin{array}{c} n \\ i \end{array} \right) 2^i \; 2^{n-i} \quad 2^{n^2-2n}$$

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 $\left(\begin{array}{c} n \\ \frac{49n}{100} \end{array} \right)$. 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.

$$\frac{\left(\begin{array}{c} n \\ 2 \end{array} \right) 2 \cdot \left(\begin{array}{c} n \\ \frac{49n}{100} \end{array} \right) 2^n \quad 2^{n^2-2n}}{2^{n^2}}$$

which is

$$\frac{\left(\begin{array}{c} n \\ 2 \end{array} \right) \left(\begin{array}{c} n \\ \frac{49n}{100} \end{array} \right)}{2^{n-1}}$$

For the $\left(\begin{array}{c} n \\ \frac{49n}{100} \end{array} \right)$ we use Stirling's approximation formula which finally yields (after upperbounding the square root that appears in the denominator) the following expression

$$\frac{\left(\begin{array}{c} n \\ 2 \end{array} \right) 2^{\alpha n}}{2^{n-1}}$$

where $a = -\frac{49}{100} \log \frac{49}{100} - \frac{51}{100} \log \frac{51}{100}$. 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 $\left(\begin{array}{c} n \\ \frac{49n}{100} \end{array} \right)$ term (since $b^{bn} = 2^{b \log b \cdot n}$).

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 176.

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:

$$E(\prod_{i=1}^{n} a_{i\sigma(i)}) = 1 \cdot Pr(\prod_{i=1}^{n} a_{i\sigma(i)} = 1) + 0 \cdot Pr(\prod_{i=1}^{n} a_{i\sigma(i)} = 0) = 1 \cdot 2^{-n} = 2^{-n} \tag{16.2}$$

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, using 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.

■

16.6 Puzzle

Exercise 177.

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. ■

Exercise 178.

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 $r \geq 2^{n-1}$. ■

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] B. Béla Bollobás. Random Graphs. Academic Press 1985.

[4] 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.

[5] V.Chvátal. The tail of the hypergeometric distribution. Discrete Mathematics. 25(3):285-287, 1979, Elsevier.

[6] T. C. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Introduction to algorithms. 3rd edition. MIT Press.

[7] H. Cramér. Sus un nouveau thèoréme-limite de la théorie des probabilités. Actualités Scientifiques et Industrielles. No 736. Paris 1938.

[8] P. Erdos and J. Spencer. Probabilistic methods in combinatorics. Academic Press. New York, 1974.

[9] W. Feller. An introduction to probability theory and its applications. Vol. 1. 3rd Edition. John Wiley & Sons. New York, 1968.

[10] W. Feller. An introduction to probability theory and its applications. Vol. 2. John Wiley & Sons. New York, 1971.

[11] A. V. Gerbessiotis and L. G. Valiant. Direct bulk-synchronous parallel algorithms. Journal of Parallel and Distributed Computing. 22(1994), pp. 251-267, 1994.

[12] M. T. Goodrich and R. Tamassia. Algorithm design and applications. Wiley. 2014, ISBN 978-1-118-33591-8.

[13] W. Hoeffding. Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association. 58 (301):13-30, March 1963.

[14] S. Kullback and R. A. Leibler. On information and sufficiency. The Annals of Mathematical Statistics. 22 (1):79-86, 1952.

[15] 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.

[16] C. McDiarmid . Concentration. In: Probabilistic methods for algorithmic discrete mathematics. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed (Eds), Algorithms and Combinatorics 16, Springer, 1998.

[17] National Institute of Standards and Technology. https://dlmf.nist.gov/4.5