Probability for computer science: inequalities and their bounds

Version 1.0 (June 16, 2025)

ALEXANDROS V. GERBESSIOTIS

CS DEPARTMENT
NJIT
NEWARK, NJ 07102.
Email: alexg@njit.edu

Printed on June 26, 2025

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

Part I

Review on probability

Chapter 1

Experiments and random processes

In the remainder, $\cup$ and $\cap$ are the well known set theoretic symbols for the union and intersection of two or more sets. A' or A^c denotes the complement of A that is, $A' = A^c = S - A$, where S is a reference set known as the universal set. For two sets, X and Y the difference $X - Y$ is $X - Y = X \cap Y'$. 2^X is the powerset of X the set of all possible subsets of X inclusive of of X and the empty set $\emptyset$. A set X of finite cardinality n has a powerset of cardinality 2^n.

A function $f(x)$ of real values x to real values y, is a mapping such that to every x there corresponds a value y and we describe this by writing $y = f(x)$.

The infimum of a set is the largest lower bound of a set and is the largest value that is smaller than or equal to all elements of the set.

The supremum of a set is the smallest upper bound of a set and is the smallest value that is larger than or equal to all elements of the set.

1.1 Experiments and random processes

An experiment can have a set of outcomes. Flipping or tossing a coin can result with one of two possible outcomes: H for heads or T for tails depending on whether the coin comes heads or tails respectively.

Definition 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, trial, and random process are to be used interchangeably.

Definition 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 coin experiment, $S = \{H, T\}$. Note that in bibliography Ω is used to indicate a sample space.

Definition 3 (Sample point). An element of S is known as a sample point.

That is, an outcome of an experiment is an element of the sample space S, a sample point.

Definition 4 (Event). An event A is a set of outcomes and is a subset of S.

Sometimes we refer to the outcomes of S as elementary events. Let us throw a coin three times. The sample space now is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

The sample space is of cardinality 2^n, where n is the number of repetitions of the coin experiment. An event $E_1 = \{HHT, HTH, THH\}$ indicates an event in which only one Tails was encountered in the experiment. One can also describe this event as the event where the number of Heads is an even positive integer, or an even where the number of Heads is two! There are three possible outcomes of S associated with this event. Another event $E_2 = \{HHH\}$ is a no Tails event.

Definition 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 sequence is meant to be a finite sequence or a countably infinite sequence.

1.2 Event spaces and measurable spaces

Definition 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.

Definition 7 (σ-algebra or σ-field). T is a σ-algebra or σ-field if it contains a collection of subsets of a sample space S that satisfy the following.

1. $S \in T$,
2. $A \in T$ implies that $A' = S - A \in T$ (closedness under complement), and
3. for a countable sequence $A_1, A_2, \ldots, A_i$, where $A_i \in T$ implies $\cup_i A_i \in T$ (closedness under union).

Using DeMorgan's Law ($(A \cap B)' = A' \cup B', or (A \cup B)' = A' \cap B'$), one can show also that for a countable sequence A_i, $A_i \in T$ implies $\cap_i A_i \in T$ (closedness under intersection).

Definition 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. In the remainder, we are going to stay away from too much unnecessary mathematical theory.

Definition 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. 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 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' 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.

Example 3. For a sample space S of a measurable space (S,T), and $s \in S$, we call $\{s\}$ an elementary event. 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).

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

1.3 Probability spaces

A probability measure denoted by P is a measure were $\mathbb{R}$ is replaced by the closed interval $[0,1]$ and has one additional property: $P(S) = 1$.

Definition 11 (Probability measure P). A probability measure P is a function on T defined as $P : T \mapsto [0,1]$ that satisfies the following.

1. $P(A) \geq P(\emptyset) = 0$ for $\forall A \in T$,
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. $P(S) = 1$.

In case (2) above we say that P is countably additive. We use $P(A) = \sum_{s \in A} P(s)$, where $P(s) \geq 0$.

A probability space is the triplet (S,T,P) as defined earlier in this and previous sections. Those definitions are summarized below.

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 implicitly given.

Definition 12 (Probability space). A probability space is a triplet (S,T,P), where

- S is a sample space that is, a set of outcomes,
- $T \subseteq 2^S$ is a σ-algebra on S, that is, a collection of subsets containing S and closed under complement, closed under union (of a countable number of sets), and by DeMorgan implication closed under intersection (of a countable number of sets), and
- P is a probability measure on T 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 elementary events of A.

If S is finite, (S,T,P) is a finite probability space. If S is finite and $T = 2^S$ the probability measure is determined by its values on elementary events. Thus the probability space assigns through $p : S \to [0,1]$ a probability $p(s)$ to every element of s of S such that $p(s) \geq 0$, with $\sum_{s \in S} p(s) = 1$. Then for an event $A \in 2^S$, the probability of the event A is the sum of the probabilities of the elements of S in A i.e. the sum of the probabilities of the elementary events, or in other words $P(A) = \sum_{s\in A} p(s)$.

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

Definition 13 (Discrete probability space). A discrete probability space has S that is a countable set. Then $P(A) = \sum_{s\in A} p(s)$, where $p(s) \geq 0$, and $\sum_{s\in S} p(s) = 1$.

Sometimes we refer to it as finite probability space. 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 14 (Equiprobable space). For a finite probability space S of n sample points, if each sample point has the same probability as any other one, the sample space is called equiprobable space.

1.4 Probability of events

Lemma 1. *For any collection of events $A_1,\ldots,A_n$,*

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \ldots + P(A_n) = \sum_{i=1}^{n} P(A_i).$$

Proof. Consider, events B_i such that

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

Then $\cup_i B_i = \cup A_i$ and $P(B_i) \leq P(A_i)$ and the events B_i are disjoint. By additivity of the probability measure we have

$$P(A_1 \cup \ldots \cup A_n) = P(B_1 \cup \ldots \cup B_n) = \sum_i P(B_i) \leq \sum_i P(A_i)$$

■

Lemma 2 (Independent Events). *Two events A and B are independent if*

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

Theorem 1.2 (Properties of events). *Consider two events A, B of probability space (S, T, P).*

$$P(\emptyset) = 0,$$

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

and for $A \subseteq B$ we have

$$P(A) \leq P(B).$$

Moreover,

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

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 1. Given two events A, B we have

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

Theorem 1.3 (Inclusion-Exclusion of three events)**.** *Given three events A_1, A_2, A_3 we have the following*

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

Example 4. 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. 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, 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 or T. Event Y is $Y = \{HH, HT, TH\}$ i.e. at least one H.

Example 5. An experiment is performed by tossing a coin. The experiment is repeated until an H is encountered. The sample space is infinite. Why? Because $S = \{T, TH, TTH, TTTH, TTTTH, \ldots\}$.

The singular form of dice is die.

Example 6. 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 S has 36 outcomes

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

Example 7. 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 8. A coin is tossed twice. The number of heads is recorded. The sample space is $S = \{0,1,2\}$. The following probability model is assigned $p(0) = 1/4, p(1) = 1/2, p(2) = 1/4$. Event $A = \{1,2\}$ with $p(A) = 3/4$, and event $B = \{2\}$ with $p(B) = 1/4$,

Example 9. 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 10 (Uniform distribution). $P(A) = c(A)/c(S)$ for all $A \subseteq S$. Note $c(A)$ is the cardinality of set A also denoted $|A|$.

1.5 Conditional probabilities

Theorem 1.4 (Conditional Probability). *Let A,B be two events of a finite probability space (S,T,P) and $P(B) > 0$. The probability that an event A occurs conditional on event B had already occurred is known as the conditional probability of A given B and denoted $P(A|B)$. It is given by*

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

Moreover

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

Theorem 1.5 (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 1.6. *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 1.7. *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 1.8. $P(A_1/A_2) = \frac{P(A_2/A_1)P(A_1)}{P(A_2)}$.

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

Definition 15 (Properties of independent events). Let A,B be two events of finite probability space (S,T,P). A is independent of B if

$$\text{A is independent of B} \Leftrightarrow P(A) = P(A|B).$$

Corollary 2. 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 11. A (pair of) dice is tossed. The probability of any elementary event (outcome) of S is $1/36$. Let B be the event the sum of the dice is 6. What is $P(B)$? Let A be the event a dice is 5. What is $P(A|B)$? This reads probability the event A occurs given that B has already occurred. This is because $P(A \cap B) = 2/36$. This is because $P(B) = 5/36$. And $P(A|B) = P(A \cap B)/P(B) = (2/36)/(5/36) = 2/5$.

Proof. We show B below.

$$B = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

The probability that a dice is 5 if the sum is 6 is 2/5. Then $A|B$

$$A|B = \{(1,5),(5,1)\}$$

Obviously $P(A|B) = 2/5$. This is because $P(A \cap B) = 2/36$, and $P(B) = 5/36$. And $P(A|B) = P(A \cap B)/P(B) = (2/36)/(5/36) = 2/5$. ■

Theorem 1.9. *For any event A, we have*

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

Proof.

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

■

Definition 16 (Independent Repeated Experiments). Let (S,T,P) be a finite probability space. A space of n independent trials is space S_n consisting of ordered n-tuples of elements of S with the probability of an n-tuple defined to be the product of the probabilities of its components.

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

The sample space S of an experiment or trial or process has outcomes that might be number or might not be numbers. Think about the experiment of tossing a coin. Sometimes instead of using symbolic values or names to an outcome such as H or T we prefer to use numeric values such as 1 and 0 respectively. This mapping is known as random variable.

Definition 17. 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 1.10 (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

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

Chapter 2

Random variables

2.1 Random variables

As we mentioned earlier, a probability space is a triplet (S,T,P). In the simplest of the cases, S is a finite set, and $T = 2^S = P(S)$, that is, T is the powerset of S, the set of all posible subsets of S and thus T has cardinality $|T| = 2^{|S|}$. The probability measure P, $P : T \mapsto [0,1]$, essentially becomes, $P : S \mapsto [0,1]$, i.e. it is defined on the elementary events, and by extension to any event subset of T. Then

$$P(A) = \sum_{s \in A} P(\{s\}),$$

and

$$\sum_{s \in S} P(\{s\}) = 1.$$

We can then define arbitrary functions on events.

Definition 18 (Random variable: simple definition)**.** A random variable (later, r.v.) X defined on a probability space (S,T,P) is a real-valued measurable function on S. Random variable X is then a mapping

$$X : S \mapsto \mathbb{R}.$$

We can add few more details on this definition.

Definition 19 (Random variable: more detailed definition)**.** A random variable (later, r.v.) X defined on a probability space (S,T,P) is a real-valued measurable function that assigns a real value $X(s)$ to an elementary event s of S so that for every x, where $-\infty < x < \infty$, set $\{s : X(s) \leq x\}$ is an event contained in T.

For a random variable we can informally write $P(X = 5)$ to actually mean $P(\{s \in S : X(s) = 5\}$. 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 20 (Convention). 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,

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

A random variable and can be discrete or continuous.

2.1.1 Discrete random variables

The discussion below applies to a discrete random variable.

Definition 21 (Discrete r.v.)**.** A random variable is discrete if its domain consists of a finite or countably infinite set of outcomes.

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 22 (Probability mass function)**.** A probability mass function (p.m.f.) of a discrete r.v. X is the function f_X that associates a probability to each $s \in S$.

In other words the p.m.f. of r.v. X is a function that returns $P(X = x)$ for each s in the domain of X with $X(s) = x$. Then $f_X(x) = f(x) = P(X = x) = P(E) = \sum_{s\in E} P(s)$. It is $f_X(x) \geq 0$ and $\sum_{s\in S} f_X(s) = 1$.

For a discrete random variable X that also takes discrete values $x_1, \ldots, x_n$, we have that

$$P(X = x_i) = P(\{s \in S : X(s) = x_i\}) = f(x_i),$$

Theorem 2.1 (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. For function X its range is $R(S,X)$.*

Example 12. 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

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

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. 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

$$Y : S \to \mathbb{R} \quad \text{where } Y((a,b)) = \min(a,b).$$

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. The range of Y is $R(S,Y) = \{1,2,3,4,5,6\}$.

Example 13. We have two dice. The sample space of throwing 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.

Proof. 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

$$P(X\in A)=P(\{(1,3),(2,2),(3,1)\})=3/36=1/12.$$

■

2.1.2 Continuous random variables

The discussion below applies to a continuous random variable.

Definition 23 (Continuous r.v.)**.** A random variable is continuous if its domain is uncountably infinite.

Definition 24 (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 2.2. *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 25 (Probability density function)**.** When the distribution function $F_X(x) = P(X \le x)$ has the form

$$P(X \le x) = \int_{-\infty}^{x} f_X(t)dt,$$

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 $F(x) = \sum_{t=X(s)\le x, s\in S} f(t)$.

2.1.3 Composition of random variables

Definition 26 (**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.

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

2.2 Expectation of a random variable

The term expectation or expected value is associated with a random variable. The terms mean and average can be used interchangeably with (mathematical) expectation, but the term mean usually refers to a distribution. The expectation of a random variable X would be denoted by $E(X)$ though it is more often to encounter $E[X]$. (Do not be surprised if by accident you see both instances in the remainder.) The term for mean is usually μ or $\bar{X}$.

For the discrete case we have.

Definition 27 (Expectation of a discrete r.v. X). Let X be a discrete random variable on probability space (S,T,P) that also takes values $x_1,\ldots$ with $P(X=x_i)=f_X(i)$. Then the expectation (or expected value) of r.v. X is defined as follows.

$$E(X)=\mu=\sum_{x\in X} x f_X(x)=\sum_i x_i f_X(x_i).$$

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

Bear in mind that dropping subscripts or introducing subscripts, $P_X(X=x_i)=P(X=x_i)=f(x_i)=f_X(i)$. If it does not cause ambiguity, we will be using f for f_X or P for P_X.

Definition 28 (Expectation of a continuous r.v. X). 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.

$$E(X)=\int_{-\infty}^{\infty} x f_X(x) d(x).$$

If $E(X)=-\infty$ or $E(X)=\infty$ we say the expectation $E(X)$ does not exist.

Theorem 2.3. *Let X,Y be r.v. and $a,b\in\mathbb{R}$. Then*

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

Proof. Let X, Y be two discrete random variables taking also discrete values $x_1,\ldots,x_i,\ldots$ and $y_1,\ldots,y_j,\ldots$. In the discrete case, we have the following.

$$\begin{aligned}
E(aX+bY) &= \sum_i\sum_j (ax_i+by_j)P(X=x_i\cap Y=y_j)\\
&= \sum_i\sum_j ax_iP(X=x_i\cap Y=y_j)+\sum_i\sum_j by_jP(X=x_i\cap Y=y_j)\\
&= a\sum_i x_i\sum_j P(X=x_i\cap Y=y_j)+b\sum_j y_j\sum_i P(Y=y_j\cap X=x_i)\\
&= a\sum_i x_iP(X=x_i)+b\sum_j y_jP(Y=y_j)\\
&= E(X)+E(Y).
\end{aligned}$$

Similarly, the corresponding continuous case bound can be proven. ■

Corollary 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 applied for any probability space.

Theorem 2.4. *For any random variables $X_1, \dots, X_n$ 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. ∎

2.3 Independent random variables

We can describe independence for the case of a discrete and continuous variable as follows.

Definition 29 (Independent r.v.)**.** Let X, Y be two discrete random variables taking also discrete values $x_1, \dots, x_i, \dots$ and $y_1, \dots, y_j, \dots$. The two random variables X and Y are independent if for all x_i and y_j,

$$P(X = x_i \cap Y = y_j) = P(X = x_i) \cdot P(Y = y_j).$$

In the general case,

$$P(X \in A \cap Y \in B) = P(X \in A) \cdot P(Y \in B).$$

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

Definition 30 (Independent r.v.: continuous case)**.** Let X and Y be two random variables with $F_X(x)$ and $F_Y(y)$ distribution functions respectively. Let $f_{X,Y}(x,y)$ be their joint distribution function. Then X and Y are independent r.v. if, for all x, y,

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

Theorem 2.5. *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 14. From a prior example of dice tossing. Random variable Y is defined as follows

$$Y : S \to \mathbb{R} \quad \textbf{where } Y((a,b)) = \min(a,b).$$

Find $E(Y)$. 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

$$\mu = E(Y) = 1 \cdot 11/36 + 2 \cdot 9/36 + 3 \cdot 7/36 + 4 \cdot 5/36 + 5 \cdot 3/36 + 6 \cdot 1/36 = \frac{11 + 18 + 21 + 20 + 15 + 6}{36} = 2.5277.$$

2.4 Indicator random variable

Definition 31 (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 2.6. *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 15. 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

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

Example 16. 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.)

2.5 More on expectation

Definition 32 (Conditional expectation: discrete case)**.** For two discrete random variables X and Y, the conditional expectation of X given Y is

$$E(X|Y = y) = \sum_{x \in X} xP(X = x|Y = y) = \sum_{x \in X} x \frac{P(X = x, Y = y)}{P(Y = y)},$$

where $P(X = x, Y = y)$ is the joint probability (mass) function of X and Y. An alternative formulation is as follows.

$$E(X|Y = y_j) = \sum_{x_i \in X} xP(X = x_i|Y = y_j) = \sum_{x_i \in X} x_i \cdot \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)},$$

Definition 33 (Conditional expectation: continuous case)**.** For two continuous random variables X and Y, with joint density $f_{X,Y}(x,y)$ the conditional expectation of X given Y is given below. Note that $f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y)$.

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

2.6 Variance of random variables

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 34 (**Variance of r.v. X**)**.** Let X be a random variable on a probability space S i.e. (S,T,P). Then the variance $\sigma^2(X) = var(X)$ of r.v. X is defined as follows.

$$\sigma^2(X) = var(X) = E((X-\mu)^2) = E((X-E(X))^2).$$

The standard deviation of X is defined as $\sqrt{var(X)} = \sigma(X)$.

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

Corollary 4. The variance $var(X)$ of a random variable X we have

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

Proof.

$$var(X) = E((X-E(X))^2) = E(X^2 - 2XE(X) + (E(X))^2) = E(X^2) - 2(E(X))^2 + (E(X))^2) = E(X^2) - (E(X))^2.$$

Another way to prove this for a discrete r.v. X is as follows.

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

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

Corollary 5. Let X be a r.v. and for any $a, b \in \mathbb{R}$, we have the following.

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

Definition 35. Let X and Y be r.v. then, the covariance of X and Y is defined as follows.

$$cov(X,Y) = E((X-E(X)) \cdot (Y-E(Y))) = E(XY) - E(X)E(Y).$$

Corollary 6. For r.v. X, Y, we have the following.

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y).$$

Proof.

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)2 - 2E(X)E(Y) \\ &= var(X) + var(Y) + 2(E(XY) - E(X)E(Y) \\ &= var(X) + var(Y) + 2cov(X,Y). \end{aligned}$$

■

If X and Y are indepdendent $E(XY) = E(X)E(Y)$ and thus $cov(X,Y) = 0$. We then obtain the following.

Corollary 7. Let X, Y be r.v. 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.

Example 17. 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

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

■

Chapter 3

Standard distributions

3.1 Standard distributions

3.1.1 Bernoulli distribution

Definition 36 (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 a 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 8. 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.

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

■

3.1.2 Geometric distribution

Definition 37. A geometric distribution can be described as the number of Bernoullis trials to get the first success, or equivalently the number of failures before the first success. If random variable X describes the number of Bernoulli trials and variable Y the number of failure to the first success, then $X = Y + 1$. We then say $X \sim g(p)$, where p is the probability of success in a Bernoulli trial.

Corollary 9. 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 from sequences $\sum_{i=0}^{\infty} x^i = 1/(1-x)$ for $x < 1$. Then $\sum_{i=1}^{\infty} ix^{i-1} = 1/(1-x)^2$ for $x < 1$, that was used before for $E(X)$. Then $\sum_{i=2}^{\infty} i(i-1)x^{i-2} = 2/(1-x)^3$ for $x < 1$.

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

Then

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

■

3.1.3 Binomial distribution

Definition 38 (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 3.1. *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 10. 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 11. 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 3.2. *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.

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

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 18. 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.

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

■

Chapter 4

Moment generating function

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

4.1 Moment generating function

Definition 39. 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.

Example 19. 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 12. 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 4.1. *Let X_i be independent random variables, with MGF $M_{X_i}(t)$ respectively, $i = 1, \dots, n$. Let $S_n = \sum_{i=1}^n X_i$. Then the following applies for the MGF of random variable S_n.*

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

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

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

Example 20. The Uniform distribution $U(a, b)$ has

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

Example 21. The binomial distribution $B(n, p)$ has

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

Example 22. The normal distribution $N(\mu, \sigma^2)$ has

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

Example 23. The Poisson distribution $Poisson(\lambda)$ has

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

Example 24. 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 $e^t(1-p) < 1$. ■

Part II

Inequalities

Chapter 5

Probability inequalities

5.1 Probability inequalities

5.1.1 Convexity

Definition 40 (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 $f''(x) \geq 0$.

5.1.2 Markov's inequality

Theorem 5.1 (Markov's inequality). *For a non-negative random variable X with expectation μ that is thus non negative and every positive $a > 0$ we have*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof.
(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 $P(X > a) \leq E(X)/a$. ■

Another proof considers an Indicator function, and works as follows. Let

$$I = \left\{ \begin{array}{cc} 0 & X < a \\ 1 & X \geq a \end{array} \right\}$$

Then

$$P(X \geq a) = E[I] \leq E(\frac{X}{a}) \leq \frac{1}{a} E(X).$$

5.1.3 Cauchy-Schwartz inequality

Theorem 5.2 (Cauchy-Schwartz inequality). *Let A and B be two random variables. with $E(A^2) < \infty$ and $E(B^2) < \infty$. Then*

$$|E(AB)| \leq \sqrt{E(A^2)}\sqrt{E(B^2)}.$$

Proof. Consider $E((A - tB)^2) \geq 0$. Then

$$E((A - tB)^2) = E(A^2) + t^2 E(B^2) - 2tE(A)E(B) \geq 0.$$

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

$$\begin{aligned} E(A^2) + \left(\frac{E(AB)}{E(B^2)}\right)^2 E(B^2) - 2\frac{E(AB)}{E(B^2)}E(AB) &\geq 0 \\ E(A^2) + \frac{E^2(AB)}{E(B^2)} - 2\frac{E^2(AB)}{E(B^2)} &\geq 0 \\ E(A^2) - \frac{E^2(AB)}{E(B^2)} &\geq 0 \\ E(A^2)E(B^2) &\geq E^2(AB). \end{aligned}$$

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

5.1.4 Chebyshev inequality

Theorem 5.3 (Chebyshev's inequality). *Let X be a random variable X with finite expected value μ and non-zero and finite standard deviation σ (variance σ^2). Then for every positive $t > 0$,*

$$P(|X - \mu| \geq \sigma t) \leq \frac{1}{t^2}.$$

Proof. It derives directly from Markov's inequality. Note that in Markov's case μ must be non-negative. This is not needed here since the use of $(X-\mu)^2$ implies an expected value that is non-negative.

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

∎

If one replaces σt with t the following alternative form is derived.

Corollary 13 (Chebyshev variation 1)**.** An alternative form is the following bound.

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

The following form starts from $P(|X| \geq t)$ instead of $P(|X-\mu| \geq t)$ or $P(|X-\mu| \geq \sigma t)$.

Corollary 14 (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 $k > 2$.

Example 25. Let S 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 $\sigma^2 = (1/n)\sum_i var(X_i)$.

If we repeat an experiment n times, we can consider the outcome of each experiment a random variable and thus describe the n experiments with r.v. $X_1, X_2, \ldots, X_n$, where X_i is the random variable associated with the i-the outcome. If the experiments are independent of each other, then $P(X_i = x_k, X_j = x_l) = P(X_i = x_k)P(X_j = x_l)$.

Proof. For S_n as defined, we have $E[S_n] = \sum_i E[X_i]$.

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

■

Example 26. 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,

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

■

5.1.5 Jensen's inequality

Theorem 5.4 (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 $g(c_1x_1 + c_2x_2) \leq c_1 g(x_1) + c_2 g(x_2)$ and this generalizes for $n > 2$, so that for $c_1 + c_2 + \ldots + c_n = 1$, we have $g(c_1x_1 + \ldots c_nx_n) \leq c_1 g(x_1) + \ldots + c_n g(x_n)$. 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

$$E(g(X)) \geq (f(x_1) + f(x_2))g\left(\frac{f(x_1)}{(f(x_1) + f(x_2)}x_1 + \frac{f(x_2)}{(f(x_1) + f(x_2)}x_2\right) + \sum_{i=3}^{n} f(x_i)g(x_i). \tag{5.1}$$

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.

■

5.1.6 Law of large numbers

Definition 41 (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 42 (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.

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

5.1.7 Tails of the binomial distribution

The following result in the form of Equations (5.3) and (5.4) appears in Feller [9] and in its complete form including Equation (5.2) as Theorem 1.1 in Bóllobás [3], where $B(n,p;m)$ is expanded and bounded,

Theorem 5.5 (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 $m = \lceil upn \rceil$, $u > 1$,*

$$P(S_n \geq m) \leq \frac{u}{u-1} B(n,p;m). \tag{5.2}$$

Furthermore,

$$P(S_n \geq m) \leq \frac{m(1-p)}{(m-np)^2}. \tag{5.3}$$

and

$$P(S_n \leq m) \leq \frac{(n-m)p}{(np-m)^2}. \tag{5.4}$$

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

$$\frac{B(n,p;k)}{B(n,p;k-1} = 1 + \frac{(n+1)p-k}{k(1-p)}. \tag{5.5}$$

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

$$\frac{B(n,p;k+1)}{B(n,p;k} = \frac{(n-k)p}{(k+1)(1-p)}. \tag{5.6}$$

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. (5.6) 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

$$P(S_n \geq m) \leq \frac{1}{1-r} B(n,p;m). \tag{5.7}$$

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

$$\begin{aligned} r &= \frac{(n-m)p}{m(1-p)} \Rightarrow \\ \frac{1}{1-r} &= \frac{1}{1-\frac{(n-m)p}{m(1-p)}} \\ &= \frac{m(1-p)}{m-np} \end{aligned} \tag{5.8}$$

Therefore Eq. (5.7) by way of Eq. (5.8) becomes.

$$P(S_n \geq m) \leq \frac{1}{1-r}B(n,p;m) \leq B(n,p;m)\frac{m(1-p)}{m-np} \tag{5.9}$$

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

$$\begin{aligned}(m-np)B(n,p;m) &\leq \sum_{i=M}^{m} B(n,p;i) \leq 1 \\ B(n,p;m) &\leq \frac{1}{(m-np)}.\end{aligned} \tag{5.10}$$

Therefore we conclude that Eq. (5.9) by way of Eq. (5.10) is as follows.

$$P(S_n \geq m) \leq B(n,p;m)\frac{m(1-p)}{m-np} \leq \frac{m(1-p)}{(m-np)^2}, \tag{5.11}$$

as needed in Eq. (5.3). We derive Eq. (5.4) by using the antisymmetry of the result of Eq. (5.3). In Eq. (5.4) we are interested in at most m successes therefore at least $n-m$ failures. Therefore we rewrite the just derived Eq. (5.3), 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.

$$P(S_n \geq n-m) \leq \frac{(n-m)p}{((n-m)-n(1-p))^2} \leq \frac{(n-m)p}{(mp-n)^2}. \tag{5.12}$$

This is a bound for failures $\geq n-m$ that also translates to a bound for successes $\leq m$ and this is Eq. 5.4, as required. For Eq. 5.2 consider $r<1/u$ and thus in the derivation of Eq.5.8 $1/(1-r)<u/(u-1)$, as needed. ■

5.1.8 Example

Example 27. 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. ■

Chapter 6

Chernoff's inequalities

6.1 Chernoff's method

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 6.1 after some relevant definitions. One can extract a variety of bounds out of Theorem 6.1. Such bounds can lead to a sequence of lemmas such as Lemma 3, Lemma 4, Lemma 5, and Lemma 6 for the right tails, and Lemma 7, and Lemma 8, and Lemma 9 for the left tails, and their associated corollaries along with some obvious concentration bounds (left and right tail bounds).

Definition 43. 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 $F_X(x) = P(X \le x)$. 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 $P(X \le 0) > 0$ and $P(X \ge 0) > 0$ then $m(0) > 0$. The following is Theorem 1 of Chernoff [4]

Theorem 6.1 (Chernoff [4])**.** *If $E(X) < \infty$ and $r \ge E(X)$ then*

$$P(S_n \ge nr) \le (m(r))^n. \tag{6.1}$$

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

$$P(S_n \le nr) \le (m(r))^n. \tag{6.2}$$

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

$$\lim_{n\to\infty} \frac{(m(r)-u)^n}{P(S_n \le nr)} = 0. \tag{6.3}$$

Proof. In order to prove Eq.(6.1) 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 $f(x) = e^{tx}$.

$$\begin{aligned} P(S_n \geq nr) &= P(tS_n \geq tnr) \\ &= P(e^{tS_n} \geq e^{nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{-nrt}E(e^{tS_n}) && (6.4) \\ &= e^{-nrt}M_{S_n}(t). && (6.5) \end{aligned}$$

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

$$\begin{aligned} E(e^{tS_n}) &= E(e^{t\sum_i X_i}) \\ &= E(\prod_i e^{tX_i}) \\ &= \prod_i E(e^{tX_i}) \\ &= \prod_i M_{X_i}(t) \\ &= (M_X(t))^n. && (6.6) \end{aligned}$$

From Eq. (6.5) by way of Eq. (6.6) we obtain the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt}E(e^{tS_n}) \\ &\leq e^{-nrt}\left(M_X(t)\right)^n \\ &= \left(e^{-rt}M_X(t)\right)^n \\ &= \inf_{t>0}\left(e^{-rt}M_X(t)\right)^n \\ &= \inf_{t>0}\left(m(r)\right)^n. && (6.7) \end{aligned}$$

This completes the proof of Eq.(6.1).
The proof of Eq.(6.2) is similar. Consider a $t < 0$ now.

$$\begin{aligned} P(S_n \leq nr) &= P(tS_n \geq tnr) \\ &= P(e^{tS_n} \geq e^{nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{-nrt}E(e^{tS_n}) \\ &= e^{-nrt}M_{S_n}(t). && (6.8) \end{aligned}$$

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 44 (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 $\sum_i p_i = \sum_i q_i = 1$.

$$D_{KL}(P||Q) = D(P||Q) = \sum_i p_i \ln \frac{p_i}{q_i}. \tag{6.9}$$

6.2 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.(6.1).

Lemma 3. *Let X_i be independent random variables following the same distribution as random variable X, $X \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$. Then, for any r such that $p < r < 1$ we have the following.*

$$P(S_n \geq rn) \leq \exp\left(-D(r||p)n\right) = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{6.10}$$

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 20.

Proof. (Of Eq. (6.10))
By Eq. (6.7) we have the following considering that $M_X(t) = E(e^{tX}) = 1 + p(e^t - 1) \leq e^{pe^t - p}$.

$$\begin{aligned}
P(S_n \geq nr) &\leq \inf_{t>0} \left(e^{-rt} M_X(t)\right)^n \\
&\leq \inf_{t>0} \left(e^{-rt}(1 + p(e^t - 1))\right)^n \\
&\leq \inf_{t>0} \left(\frac{(1 + p(e^t - 1))}{e^{rt}}\right)^n \qquad (6.11)\\
&\leq \inf_{t>0} \left(\exp\left(f(t)\right)\right)^n \qquad (6.12)
\end{aligned}$$

As indicated by Eq. (6.12), $f(t) = \ln\left((1 + p(e^t - 1)\right) - rt$. 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. (6.11) as follows.

$$P(S_n \geq nr) \leq \inf_{t>0} \left(\frac{(1 + p(e^t - 1))}{e^{rt}}\right)^n \tag{6.13}$$

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

$$\exp(rt) = ((1-p)r/(p(1-r)))^r = \frac{(1-p)^r r^r}{p^r (1-r)^r}. \tag{6.14}$$

Likewise the numerator is as follows.

$$(1 + p(e^t - 1)) = \frac{(1-p)rp}{p(1-r)} + 1 - p = \frac{1-p}{1-r} \tag{6.15}$$

Therefore we have the following for the quantity below.

$$\frac{(1+p(e^t-1))}{e^{rt}} = \frac{1-p}{1-r} \cdot \frac{p^r(1-r)^r}{(1-p)^r r^r} = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{6.16}$$

The proof is completed. ■

We generate one more bound from Eq.(6.1). The bound is stronger than the Corollaries to follow that bring it into an easier to deal with form.

Lemma 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 r such that $p < r < 1$ we have the following, after we substiture $r = (1+\delta)p$, $\delta > 0$.*

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} \tag{6.17}$$

Proof. By Eq. (6.7) we use $M_X(t) = E(e^{tX}) = 1 + p(e^t-1) \leq e^{pe^t - p}$, to obtain the following.

$$\begin{aligned} P(S_n \geq nr) &\leq \inf_{t>0} \left(e^{-rt}(1+p(e^t-1))\right)^n \\ &\leq \inf_{t>0} \left(e^{-rt} e^{pe^t - p}\right)^n && (6.18) \\ &\leq \inf_{t>0} \left(\exp\left(pe^t - p - rt\right)\right)^n \\ &\leq \inf_{t>0} \left(\exp\left(f(t)\right)\right)^n && (6.19) \end{aligned}$$

The second to last part is because for all x, we have $e^x \geq 1+x$. By Eq. (6.18) and Eq. (6.19) 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.(6.19) the following is obtained.

$$\begin{aligned} P(S_n \geq nr) &\leq \inf_{t>0} \left(\exp\left(pe^t - p - rt\right)\right)^n \\ &\leq \left(\exp\left(p(r/p) - p - r\ln(r/p)\right)\right)^n \\ &\leq \left(\frac{e^r \cdot p^r}{e^p \cdot r^r}\right)^n && (6.20) \end{aligned}$$

Finally we substitute $r = (1+\delta)p$ in Eq. (6.20) to obtain our result

$$\begin{aligned} P(S_n \geq nr) &\leq \left(\frac{e^{(1+\delta)p} \cdot p^{(1+\delta)p}}{e^p \cdot ((1+\delta)p)^{(1+\delta)p}}\right)^n \\ &\leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{pn} && (6.21) \end{aligned}$$

■

A simpler proof of Eq. (6.10) for binomial random variables is found in [5] and the proof is presented below.

Lemma 5 ([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.*

$$P(S_n \geq rn) \leq \exp\left(-D(r||p)n\right) = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{6.22}$$

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

$$\begin{aligned} B(n,p,k) &\leq \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^{i-k} \\ &\leq x^{-k} \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^i \\ &\leq x^{-k} \sum_{i=0}^{n} \binom{n}{i} (px)^i (1-p)^{n-i} \\ &\leq x^{-k} (1 + (x-1)p)^n. \end{aligned} \tag{6.23}$$

For $r > p$ consider $x = (1-p)r/(p(1-r))$ and substitute for the x of Eq. (6.23). We obtain the following.

$$\begin{aligned} B(n,p,k) &\leq x^{-k}(1+(x-1)p)^n \\ &\leq \left(\frac{(1-p)r}{p(1-r)}\right)^{-rn} \cdot \left(1 + \left(\frac{(1-p)r}{p(1-r)} - 1\right)p\right)^n \end{aligned} \tag{6.24}$$

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 6. *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.*

$$P(S_n \geq rn) \leq \exp\left(-2n(r-p)^2\right). \tag{6.25}$$

Proof. The proof follows the steps of the proof of Theorem 6.1 to Eq.(6.4).

$$P(S_n \geq nr) = P(tS_n \geq tnr) \leq \frac{E(e^{tS_n})}{e^{nrt}} \leq e^{-nrt} E(e^{tS_n}) \tag{6.26}$$

We then proceed differently as follows.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt} E(e^{tS_n}) \\ &\leq e^{-nrt} e^{npt} E(e^{t(S_n - np)}) \\ &\leq e^{-nrt+npt} E(\prod_i e^{t(X_i - p)}) \\ &\leq e^{-nrt+npt} \prod_i E(e^{t(X_i - p)}) \end{aligned} \tag{6.27}$$

We note that r.v. $X_i - p$ is bounded and $E(X_i - p) = E(X_i) - p = p - p = 0$. Then, Proposition 10 that is being utilized in Hoeffding bounds can be used to show the following.

$$E(e^{t(X_i-p)}) \leq \exp\left(\frac{t^2}{8}\right). \tag{6.28}$$

Eq.(6.27) by way of Eq.(6.28) yields the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt+npt} \prod_i E(e^{t(X_i-p)}) \\ &\leq e^{-nrt+npt} \prod_i \exp\left(\frac{t^2}{8}\right) \\ &\leq e^{-nrt+npt} \exp\left(n\frac{t^2}{8}\right) \\ &\leq e^{-nrt+npt+\frac{nt^2}{8}}. \end{aligned} \tag{6.29}$$

Consider the exponent of Eq.(6.29): $f(t) = -nrt + npt + nt^2/8$. 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

$$t = 4(r-p). \tag{6.30}$$

Given that $f^{''}(t) = 1/4 > 0$, we have a minimum at $t = 4(r-p)$ for $f(t)$. Therefore Eq.(6.29) by way of Eq.(6.30) yields the following.

$$\begin{aligned} P(S_n \geq nr) &\leq e^{-nrt+npt+\frac{nt^2}{8}} \\ &\leq e^{-nr\cdot 4(r-p)+np\cdot 4(r-p)+\frac{n\cdot 16(r-p)^2}{8}} \\ &\leq e^{-2n(r-p)^2}. \end{aligned} \tag{6.31}$$

Eq.(6.31) is Eq.(6.25) as needed. ■

The following corollary is more widely known than Lemma 6.

Corollary 15. 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.

$$P(S_n - E(S_n) \geq rn) \leq \exp\left(-2nr^2\right). \tag{6.32}$$

Proof. In Lemma 6 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.(6.25) to obtain Eq.(6.32). ■

The following Corollary can be obtained from Lemma 4. It provides a more tangible upper bound than the generic one of the Lemma.

Corollary 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 $\delta > 2e - 1$. we have the following,

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq 2^{-(1+\delta)pn}. \tag{6.33}$$

Proof. From Lemma 4, 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 4. Note that $\delta > 0$ is better than the δ used in Corollary 19 that follows it.

Corollary 17. Let X_i be independent random variables following the same distribution as random variable X, $X \sim b(p)$, where $0 < p < 1$. Let $S_n = \sum_i X_i$. Then, for any $\delta > 0$ we have the following.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{2+\delta} \cdot pn\right). \tag{6.34}$$

Proof. We would like to upper bound the bound of Eq.(6.17) of Lemma 4, 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.

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

By inequality 20

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

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 18 ([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.

$$P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{2+2\cdot\delta/3} \cdot pn\right). \tag{6.35}$$

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 4. It is similar to the one in [2] for the binomial case.

Corollary 19. 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.

$$P(S_n \geq rn) = P(S_n \geq (1+\delta)pn) \leq \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{6.36}$$

Proof. By inequality (27) 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 17, 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 3. This is Corollary 20.

Corollary 20. 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.

$$P(S_n \geq (p+t)n) \leq \exp(-D((p+t)||p)n) = \left[\left(\frac{p}{p+t}\right)^{p+t}\left(\frac{1-p}{1-p-t}\right)^{1-p-t}\right]^n. \tag{6.37}$$

6.3 Derived left tails

We proceed to deriving a Lemma identical to Lemma 3 for the left tails. This is stated next.

Lemma 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 r such that $0 < r < p$ we have the following.*

$$P(S_n \leq rn) \leq \exp\left(-D(r||p)n\right) = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^n. \tag{6.38}$$

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

Proof. (Of Eq. (6.38))
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 3 is bounded above by Eq.(6.10) 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

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

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

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

Method 2. Now let us reprove it following the method of the proof of Lemma 3. Consider $t > 0$, and apply the Chernoff trick and finally use Markov's inequality as we did earlier.

$$\begin{aligned} P(S_n \leq nr) &= P(tS_n \leq tnr) \\ &= P(-tS_n \geq -tnr) \\ &= P(e^{-tS_n} \geq e^{-nrt}) \\ &\leq \frac{E(e^{tS_n})}{e^{nrt}} \\ &\leq e^{nrt} E(e^{-tS_n}) \\ &\leq e^{nrt} (E(e^{-Xt}))^n \end{aligned} \tag{6.39}$$

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

$$\begin{aligned} P(S_n \leq nr) &\leq \inf_{t>0} \left(e^{rt}(1 + p(e^{-t} - 1))\right)^n \\ &\leq \inf_{t>0} \left(\exp\left(f(t)\right)\right)^n. \end{aligned} \tag{6.40}$$

As indicated by Eq. (6.40), $f(t) = \ln\left(1 + p(e^{-t} - 1)\right) + rt$. 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. (6.40) as follows.

$$P(S_n \leq nr) \leq \inf_{t>0} \left(e^{rt}(1 + p(e^{-t} - 1))\right)^n \tag{6.41}$$

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

$$\exp(rt) = \frac{p^r(1-r)^r}{(1-p)^r r^r}. \tag{6.42}$$

Likewise the other term is as follows.

$$(1+p(e^{-t}-1)) = \frac{p(1-r) - p^2(1-r) + p(1-p)r}{p(1-r)} = \frac{1-p}{1-r} \tag{6.43}$$

Therefore we have the following for the quatinty below.

$$e^{rt} \cdot (1+p(e^{-t}-1)) = \frac{1-p}{1-r} \cdot \frac{p^r(1-r)^r}{(1-p)^r r^r} = \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r} \right]^n. \tag{6.44}$$

■

We generate one more bound below.

Lemma 8. *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.*

$$P(S_n \le rn) = P(S_n \le (1-\delta)pn) \le \left(\frac{e^r \cdot p^r}{e^p \cdot r^r} \right)^n = \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{pn} \tag{6.45}$$

Proof. We calculated earlier $E(e^{-tX}) = 1 + p(e^{-t}-1) \le e^{pe^{-t}-p}$. By way of Eq.(6.40) we have the following.

$$\begin{aligned} P(S_n \le nr) &\le \inf_{t>0} \left(e^{rt}(1+p(e^{-t}-1)) \right)^n \\ &\le \inf_{t>0} \left(\exp(pe^{-t} - p + rt) \right)^n && (6.46) \\ &\le \inf_{t>0} \left(\exp(f(t)) \right)^n && (6.47) \end{aligned}$$

By Eq. (6.46) and Eq. (6.47) 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.(6.46) the following is obtained.

$$\begin{aligned} P(S_n \le nr) &\le \inf_{t>0} \left(\exp(pe^{-t} - p + rt) \right)^n \\ &\le (\exp(pr/p - p + r\ln(p/r))^n \\ &\le \left(\frac{e^r \cdot p^r}{e^p \cdot r^r} \right)^n && (6.48) \end{aligned}$$

Finally we substitute $r = (1-\delta)p$ in Eq. (6.48) to obtain Eq.(6.45).

$$\begin{aligned} P(S_n \le nr) &\le \left(\frac{e^{(1-\delta)p} \cdot p^{(1-\delta)p}}{e^p \cdot ((1-\delta)p)^{(1-\delta)p}} \right)^n \\ &\le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{pn} && (6.49) \end{aligned}$$

■

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 21. 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.

$$P(S_n \leq (1-\delta)pn) \leq \exp\left(\frac{-\delta^2}{2} \cdot pn\right). \tag{6.50}$$

Proof. By inequality (26) 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 6 is stated below.

Lemma 9. *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.*

$$P(S_n \leq rn) \leq \exp\left(-2n(r-p)^2\right). \tag{6.51}$$

Proof. The proof is by symmetry to Lemma 6 for $Y_i = 1 - X_i$ and $\sum_i Y_i = n - S_n$ instead. ■

The following corollary is then evident.

Corollary 22. 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.

$$P(S_n - E(S_n) \leq rn) \leq \exp\left(-2nr^2\right). \tag{6.52}$$

6.4 Derived concentration bounds

Finally the following Corollary can also be obtained from Corollary 19 and Corollary 21.

Corollary 23. 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,

$$P(|S_n - np| \geq \delta pn) \leq 2 \cdot \exp\left(\frac{-\delta^2}{3} \cdot pn\right). \tag{6.53}$$

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 (19) we bound $P(S_n \geq (1+\delta)pn)$. By Corollary 21 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 15 and Corollary 22.

Corollary 24. 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.

$$P(|S_n - E(S_n)| \geq rn) \leq 2 \cdot \exp\left(-2n(r-p)^2\right). \tag{6.54}$$

Chapter 7

Hoeffding's inequalities

7.1 Hoeffding's method

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. $0 \le X_i \le 1$ or $a_i \le X_i \le b_i$. Theorem 1 [13] for the case of Bernoulli trials takes the form of Corollary 20 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.*

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

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.

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

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

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

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.

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

Therefore,

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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 $\exp\left(-nh^2 k(p)\right)$, 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 $\exp\left(-2nh^2\right)$ 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 10 (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 $f(x) = \exp\left(tx\right)$ 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

$$E(e^{tX}) \leq \exp\left(\lg\left(\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}\right)\right) = \exp\left(\ln\left(g(t)\right)\right).$$

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,

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

■

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 10 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.

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

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

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.*

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

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 25. 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.

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

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 25 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 26 appears in [15] as 5.3 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 26. 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.

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

Proof. Replace hn in Theorem 7.2 with h. ■

Moreover one can combine Corollary 26 and Corollary 30 to bound $P(|S_n - pn| \geq h)$.

The following is due to [2]. It also appears as Corollary 19 and Corollary 21

Theorem 7.4 (Angluin-Valiant[2]). *For every n, p, b with $0 \le p \le 1$ and $0 \le b \le 1$, we have the following.*

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

and

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

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

Corollary 27. Let X_i be an independent random variable, $i = 1,\dots,n$, 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.

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

Corollary 28. Let X_i be an independent random variable, $i = 1,\dots,n$, 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.

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

7.3 Derived left tails

Corollary 29. Let X_i be an independent random variable, $i = 1,\dots,n$, such that $a \le X_i \le b$. 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.

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

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

Corollary 30 appears in [15] as 5.4 in Corollary 5.2 for $a = 0$ and $b = 1$.

Corollary 30. Let X_i be an independent random variable, $i = 1,\dots,n$, such that $a \le X_i \le b$. 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.

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

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

Corollary 31. Let X_i be an independent random variable, $i = 1, \dots, n$, 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.

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

Corollary 32. Let X_i be an independent random variable, $i = 1, \dots, n$, such that $\le X_i \sim b(p)$, 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.

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

7.4 Derived concentration bounds

The following Corollary combines Corollary 25 with Corollary 29.

Corollary 33. Let X_i be an independent random variable, $i = 1, \dots, n$, such that $a \le X_i \le b$. 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.

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

The following Corollary combines Corollary 26 with Corollary 30 instead.

Corollary 34. Let X_i be an independent random variable, $i = 1, \dots, n$, such that $a \le X_i \le b$. 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.

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

Chapter 8

Miscellanea

8.1 Combinatorial inequalities

Inequality 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 2. For any integer $n > 0$ we have the following.

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

Inequality 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 4. For any integer $n > 0$ and $k \leq n$ we have the following.

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

Inequality 5. 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 6. 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 7 ([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.

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

8.2 Series

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

Inequality 8 (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 11 after a Taylor series expansion of $\ln(x)$ for $x = 1$, and then substitution $x+1$ for x. ■

Inequality 9. For $-1 \leq x < 1$ we have the following equality by substituting $-x$ for x in Inequality (8).

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

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 10. 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 11. 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 12. For we have as follows.

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

8.3 Exponential inequalities

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

Inequality 13. 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 14. 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 15. 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 16. 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 17. For every $x > 0, y > 0$ we have the following.

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

Inequality 18. 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 19. 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

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

■

8.4 Logarithmic and other inequalities

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

Inequality 20. 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 21. 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 22. 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 23. 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 24. For every $x \geq 0$ we have, as a consequence of Inequality 8, the following hold.

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

and

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

Inequality 25. For every $0 < x < 0.5828$ we have the following.

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

Inequality 26. For every $0 \le \delta < 1$ we have the following.

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

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

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

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

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

■

Inequality 27. For every $0 \le \delta \le 1$ we have the following.

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

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

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

We can then conclude since $0 \le x \le 1$ that

$$(1+x)\ln(1+x) \ge x + x^2/2 - x^3/6 \ge x + x^2/2 - x^2/6 \ge x + x^2/3.$$

Set $x = \delta$ and the result follows. ■

Inequality 28. For every $2e - 1 < \delta$ we have the following.

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

Inequality 29.
(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 20. (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 30 ([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.

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

8.5 Combinatorial equalities

All numbers are non-negative integers unless otherwise stated.

Lemma 11. *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 12. *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 8.1. *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 35. We have the following.

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

Corollary 36.

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

Corollary 37. 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 38. 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 13. *For any n,k,m, with $k \le r \le n$, 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 14. *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 15. *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 III

Probability problems

Chapter 9

Entropy function and random vectors

Problem 1.
Find the maximum of the entropy function $f(x) = -x\lg x - (1-x)\lg(1-x)$. $0 < x < 1$.

Proof.
We are using logarithms base two. Note that Consider $2^y = e$ i.e. $y = \lg e$ by definition. Taking $\ln$ of both sides we hage $y\ln 2 = 1$. Thus $\lg e = y = 1/\ln 2$. It is easy with a calculator to establish that $\lg e > 1$ and in fact and $\lg e \approx 1.442$. Consider now,

$$2^{\lg x} = e^{\ln x} = x.$$

Take $\lg$ of both sides. Therefore $\lg x = \ln x \lg e$. We use it to convert $f(x)$ into

$$f(x) = -x\lg x - (1-x)\lg(1-x) = -x\ln x\lg e - (1-x)\ln(1-x)\lg e = g(x)\lg e$$

where $g(x)$ is as follows.

$$g(x) = -x\ln x - (1-x)\ln(1-x).$$

Note that $f^{'}(x) = g^{'}(x)\lg e$. Find the first derivative of $g(x)$.

$$g^{'}(x) = -\ln x - x/x + \ln(1-x) - (1-x)(-1)/(1-x) = -\ln x + \ln(1-x)$$

Setting $g^{'}(x) = 0$ and solving for x we get $x = 1/2$. This is also the case for $f^{'}(x)$. We note that

$$f(1/2) = 1,$$

and that's the minimum value. For this to be the case we should have confirmed earlier that $f''(x) < 0$ or equivalently $g''(x) < 0$. Note that $f^{''}(x) = g^{''}(x)(\lg e)^2$, and the last term is always positive. The monotonicity of $f^{''}(x)$ is the monotonicity of $g^{''}(x)$. Needless

$$g^{''}(x) = -1/x - 1/(1-x)$$

and thus $g^{''}(1/2) = -4 < 0$. This means that $f(x)$ and also $g(x)$ has a maximum at $x = 1/2$. It is not difficult to establish that $f(1/2) = 1$. ∎

Problem 2.

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

Proof.
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 $x = yn = n/2$. ■

Problem 3.

You are given a random vector $a = (a_1, \dots a_n)$, 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 $0, \dots p-1$), 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.

Proof.
(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. $\sum d_i a_i \equiv 0 \bmod p$.

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

where $Z \equiv (-d_2 a_2 - \dots - d_n a_n) \bmod p$. 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 $0, \dots, p-1$. 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. ■

Chapter 10

Balls and bins, bernoulli and binomial

Problem 4.
Examine the monotonicity of the binomial process $B(n, p)$.

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

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

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 $(n+1)p - 1 < m \leq (n+1)p$. ∎

Problem 5.
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 $t > np$.

Proof.
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 $(n+1)p - 1 < m \leq (n+1)p$. ∎

Problem 6.

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.

Proof.
Consider Corollary 16. For $\delta > 2e - 1$, the following applies.

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

If we pick $(1+\delta) = 6$ the result follows. ■

Problem 7.
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?

Proof.
Use Chernoff Corollary 21

$$P(S_n \leq (1-\delta)np) \leq \exp\left(-\delta^2 np/2\right).$$

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

$$P(S_n \leq 25) \leq \exp -(1/4)25 < 2/1000.$$

■

Problem 8.

(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.

Proof.
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 $var(S_n) = npq = n(1/2)(1/2) = n/4$.
Let us go for the familiar Corollary 23. The term below (n/2) is introduced to form $np = n/2$ next to δ per Corollary 23.

$$P(|S_n/n - 1/2| \geq \delta) = P(|S_n - n/2| \geq \delta n) = P(|S_n - n/2| \geq 2\delta(n/2)) \leq 2\exp(-(n/2)4\delta^2/3) \leq 2\exp(-(2/3)n\delta^2).$$

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 (e.g. Corollary 33.

$$P(|S_n/n - 1/2| \geq \delta) = P(|S_n - n/2| \geq 2\delta(n/2)) \leq 2\exp(-2n\delta^2).$$

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 $2/n^{12}$. ∎

Problem 9.

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 $r > np$,

$$P(S_n - np \geq r) \leq \left(\frac{npe}{r}\right)^r.$$

Proof.
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.

$$e^{-\ln(r/(np))r + np(r/(np)))} = (\frac{np}{r})^r e^r = (\frac{npe}{r})^r.$$

∎

Problem 10.

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 $1 \leq r < n$.

$$P(S_n \geq r) \leq \left(\frac{npe}{r}\right)^r.$$

$$P(S_n < r) \leq \binom{n}{n-r}(1-p)^{n-r}.$$

Proof.

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 $\{1,2,\dots,n\}$. 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.

$$P(S_n \geq r) \leq P(\cup E_S) \leq \sum_{S,|S|=r} P(E_S) \leq \binom{n}{r} p^r. \tag{10.1}$$

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.(10.1) substituting failure for success and thus probability of failure $1-p$ for probability of success p. Therefore

$$P(S_n < r) = P(B(n,p;k) < r)P(B(n,(1-p);k \geq n-r) \leq \binom{n}{n-r}(1-p)^{n-r}.$$

∎

Problem 11.
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 $r \leq np$,

$$P(S_n < r) \leq \frac{r(1-p)}{np-r} B(n,p;r).$$

Proof.
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 $a/(1-a) = rq/(np-r)$. ■

Problem 12.

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 $\pi(i) \neq i \quad \forall i,\, i = 1, \ldots, n$.

Proof. 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 $n! - |A_1 \cup \ldots \cup A_n| = \sum_{i=0}^{n} (-1)^{i+1} / i!$. ■

Problem 13.

(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.

Proof.
(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

$$C(n,k) = P(n,k)/k! = n(n-1)\dots(n-(k-1)/k! = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

■

Problem 14.

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.

Proof.
(a) We describe the distribution as a n-digit sequence, each digit taking one of N values representing a bin's ID (e.g. from 1 to N). There are

$$\underbrace{N \times N \times \ldots \times \ldots N}_{\text{since there are } n \text{ digits}} = N^n$$

(b) This is equivalent to n_i the number of balls in bin i: $\sum_{i=1}^{N} n_i = n$. In other words, the number of solutions n_i of this equation in the non-negative integers is the answer to the original question. This is $\binom{n+N-1}{N-1} = \binom{n+N-1}{n}$. Think also $N+1$ vertical lines (inclusive of first and last). First and last are fixed, they can be ignored thus the position of the remaining $N-1$ varies. Total number is out of $n+N-1$ marked positions we pick $N-1$ and turn them into vertical line markers $\binom{N+n-1}{N-1}$. Between the markers we count balls per corresponding bin.
(c) Place one indistinct ball per bin. The remaining $n-N$ balls remain to be distributed into the N bins. By case (b) we have $\binom{n-1}{N-1}$. Equivalently, the n balls create $n-1$ gaps between them. Pick $k-1$ out of the $n-1$ gaps in $\binom{n-1}{N-1}$.
(d) Similar to (c). Put k_i balls as neede into bin i. The remaining balls are handled as in (b) above.

$$\binom{n - \sum_i k_i}{N-1}.$$

(e) Implied is that $N \geq n$. We need to choose which n of the N urns will be assigned a ball! Thus $\binom{N}{n}$.
(f) N choice for first ball, $N-1$ for second and so on. Total $N(N-1)\ldots(N-n+1)$. Note that if $N=n$, then the answer is $n!$. Moreover, $n > N$, answer is 0.
(g) That's a Stirling number $\left\{ \begin{matrix} N \\ n \end{matrix} \right\}$.
(h) Not many choices: if $n < N$ the answer is 1; otherwise it is zero!
(i) See the previous question. Not many choices: if $n < N$ the answer is 1; otherwise it is zero! ∎

Problem 15.

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 ?

Proof.
(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. ■

Problem 16.

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.

Proof. 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). ■

Problem 17.

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

Definition 45. 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.

Proof. 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

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

By Eq.(5.2) with $t = \lceil upn \rceil$, $p = 1/n$ and $t = \lceil u \rceil$, $u > 1$. In Eq.(5.2) the $u/(u-1)$ becomes at most 1.5 for $u > 3$ or equivalently $t > 3$. Therefore for $t > 3$

$$P(Bin\ k \geq t) = P(S_n \geq t) \leq \frac{u}{u-1} B(n, p; t) \leq 1.5 \binom{n}{t} (1/n)^t. \tag{10.2}$$

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).

$$P(Bin\ k \geq t) \leq \sum_{S_n \subseteq \{1,\ldots,n\}, |S_n|=t} \prod_i P(\text{ball i } \in S_n \text{is in bin } k) \leq \binom{n}{t} p^t = \binom{n}{t} (1/n)^t \tag{10.3}$$

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.(10.2) and Eq.(10.3) is a 1.5. We proceed discarding 1.5 as follows.

Using inequality 3 we obtain from Eq.(10.3) the following.

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

Assum $n \geq e^2$ implying a $\ln n \geq 2$. Furthermore choose $t = c\ln(n)$, with $c = e^2$. Then if $t = e^2 \ln(n)$ we have the following. Note that $2c = 2e^2 \approx 14.778 > 14$

$$\begin{aligned}
P(Bin\ k \geq t) &\leq \left(\frac{e}{t}\right)^{c\ln n} \\
&\leq \left(\frac{e}{e^2 \ln n}\right)^{c\ln n} \\
&\leq \left(\frac{1}{e\ln n}\right)^{c\ln n} \\
&\leq \left(\frac{1}{e}\right)^{c\ln n} \cdot \left(\frac{1}{\ln n}\right)^{c\ln n} \\
&\leq \frac{1}{n^c} \cdot \frac{1}{n^c} \\
&\leq \frac{1}{n^{2c}} \leq \frac{1}{n^{14}}.
\end{aligned}$$

Then

$$
\begin{aligned}
P(\text{number of balls any bin } \geq t) &= P((\text{number of balls of bin } 1 \geq t) \\
&\quad \cup(\text{number of balls of bin } 2 \geq t) \\
&\quad \cup \quad \ldots \\
&\quad \cup(\text{number of balls of bin k} \geq t) \\
&\quad \cup \quad \ldots \\
&\quad \cup(\text{number of balls of bin n} \geq t)) \\
&\leq \sum_i P((\text{number of balls of bin k} \geq t) \\
&= n \cdot P((\text{number of balls of bin k} \geq t) \\
&= n \cdot P((\text{Bin k} \geq t) \\
&\leq n \cdot \frac{1}{n^{14}} \\
&= \cdot \frac{1}{n^{13}},
\end{aligned}
$$

since $c = e^2$. This means $t = e^2 \ln(n)$ is too generous. We can make it smaller.

The bound $t = c \ln n / \ln \ln n$ for $c = 2e^2$ gives. First let's bring our bound in an exponential form that we did not do so before.

$$
\begin{aligned}
P(\text{number of balls any bin } \geq t) &\leq n \cdot P((\text{number of balls of bin k} \geq t) \\
&\leq n \cdot \frac{e^t}{t^t} \\
&\leq \exp(\ln n + t - t \ln t) \\
&\leq \frac{1}{n^6}.
\end{aligned}
$$

The bound holds for $n > 4000000 > e^{e^{e^1}}$. This is to make sure $\ln \ln \ln n / \ln \ln n$ is small enough. Thus the result has been proven w.h.p i.e. that with probability at least $1 - 1/n^6$ every bit has no more than $2e^2 \ln n / \ln \ln n = \Theta(\ln n / \ln \ln n)$ balls. ■

Problem 18.

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?

Proof.
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 $E(X) = 1 \cdot Pr(X = 1) + 0 \cdot Pr(X = 0) = (1-1/n)^{n-1}$, 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 $E(Y) = nE(X) = n\frac{n(n-1)}{2}\frac{1}{n^2}\frac{(1-1/n)^n}{(1-1/n)^2} \leq \frac{n^2}{2e(n-1)}$.

■

Problem 19.

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)!

Proof.
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.

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

■

Problem 20.
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!

Proof.
It is not difficult to conclude that

$$\frac{(n-2)^t}{n^t} = \left(1 - \frac{2}{n}\right)^t.$$

■

Problem 21.

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?

Proof.

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,

$$\begin{aligned} var(Y) = var\sum_i Y_i) &= \sum_i var(Y_i) \\ &= \sum_{i=1} n(1-p_i)/p_i^2 \\ &\leq \sum_{i=1} n1/p_i^2 \\ &= \sum_{i=1} n \frac{n^2}{(n-i+1)^2} \\ &= n^2 \sum_{i=1} n \frac{1}{(n-i+1)^2} \\ &= n^2 \sum_{i=1} n \frac{1}{i^2} \\ &= n^2 \pi^2/6, \end{aligned}$$

utilizing the Basel problem's sum due to Euler that says

$$\sum_{i=1} n \frac{1}{i^2} = \pi^2/6.$$

■

Problem 22.

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?

Proof.
(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 16. 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 4 with $r = 2/sqrtn$. The bound is an even better $\approx 1/n^{\sqrt{n}}$. Hoeffding provides the same bound through Theorem 7.1. Corollary 25 or Corollary 26 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

$$P(Y \geq 1) \leq E(Y) \leq nE(Y_i) \leq n(1/(4n)) = 1/4.$$

■

Problem 23.

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.

Proof.
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 23. 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 11

Birthdays, coupons and coins

Problem 24.

Birthday Problem

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 $1-1/2^{10}$.

Proof.
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 $(1-\frac{i}{n}) \geq e^{-2i/n}$.

$$\begin{aligned} p &= (1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})\ldots(1-\frac{m-1}{n}) \\ &\geq \exp(-2\cdot 1/n)\cdot\exp(-2\cdot 2/n)\cdot\exp(-2\cdot 3/n)\cdot\ldots\exp(-2\cdot(m-1)/n) \\ &\geq \exp(\frac{-2m(m-1)}{2n}) \geq \exp(-k^2) \end{aligned}$$

■

Problem 25.

Coupon Collector

Let $\{1,2,\ldots,n\}$ be a set of n cards (coupons). In a collection of m coupons how large should m be sort 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 $\varepsilon>0$.

Proof.
For any fixed $i\in\{1,2,\ldots,n\}$, the probability i is not chosen in m choices is given by p_i

$$p_i=\left(1-\frac{1}{n}\right)^m=\exp\left(-\frac{m}{n}\right)$$

Then the probability $p_1\vee p_2\vee p_3\vee\ldots\vee p_n$ is at most $\sum_i p_i=n\exp\left(-\frac{m}{n}\right)$. The latter probability for $m=(1+\varepsilon)n\ln n$.

$$n\exp\left(-\frac{m}{n}\right)=n\exp\left(-(1+\varepsilon)\ln n\right)=1/n^{\varepsilon}.$$

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.

∎

Problem 26.

Toss a coin $n = 2m$ times. What is the probability of exactly $n/2 = m$ heads?

Proof.
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 12

Hash tables

Problem 27.

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

$$p(n,N) \leq \exp\left(-\frac{n(n-1}{2N}\right).$$

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.

Proof.
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.

```
CP=CollisionProbability   NCP= Probability of NO collisions
                 |--> CP=0
No key-> 1stkey-                    |--> CP= 1/N
                 |-->NCP=1-->2ndkey-                        |-> CP= 2/N
                                    |-->NCP= 1-1/N-->3rdkey-
                                                            |->NCP=1-2/N
```

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. ■

Problem 28.

Hashing with chaining (continues with the following problem which is 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.)

Proof.
(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 $E[D] < k+1$. ■

Problem 29.

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?

Proof.
(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 12.1 (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 $\delta > 0$,*

$$P(X \geq (1+\delta)m) \quad \leq \quad \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^m$$

(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 $E[D] = n(1-1/n)^{n-1} \approx n^2/(e(n-1))$. ■

Problem 30.

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?

Proof.
(a) The probability that slot i is empty is the probability that all n keys missed it. In general this is $(1-1/N)^n$. Probability $1/N$ is for a key k_j to hit i or $1-1/N$ is the probability that k_j misses i. Given that we have n keys the formula derives. One can prove the first derivation below by induction on n. It also follows by expanding the power on the left side

$$(1-\frac{1}{N})^n \quad \geq \quad (1-\frac{n}{N})$$

If $N > n$ then $\geq$ becomes $>$.
(b) This is a binomial distribution's term: $b(n,k;p)$, where $k = 1$, $p = 1/N$ and thus

$$\binom{n}{1} p^k (1-p)^{n-k} \quad = \quad \frac{n}{N}(1-1/N)^{n-1}$$

Given that $N \geq n$, we have $N > n-1$. Thus from the observation and deriviation in (a) we have

$$\begin{aligned} \frac{n}{N}(1-1/N)^{n-1} \quad &> \quad \frac{n}{N}\left(1-\frac{n-1}{N}\right) \\ &= \quad \frac{n}{N} - \frac{n(n-1)}{N^2} \end{aligned}$$

(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

$$\begin{aligned} P(n_i \geq 2) \quad &\leq \quad 1-(1-\frac{n}{N}) - \frac{n}{N} + \frac{n(n-1)}{N^2} \\ P(n_i \geq 2) \quad &\leq \quad \frac{n(n-1)}{N^2} \end{aligned}$$

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 $1-1/n$. ■

Problem 31.

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 $E[T]$.

Proof.
(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 anohter problem when $x \to \infty$ and $x < 1$. We then substitute $1-p$ for x. ■

Problem 32.
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.)

Proof.
(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 $1/p_T + 1/p_H = 4$. ■

Chapter 13

Random graphs

Problem 33.

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?

Proof.
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.

■

Problem 34.

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?

Proof.
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?

∎

Problem 35.

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.

Proof.
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 $n!/2^n$.

■

Problem 36.

What is the expected number of Hamiltonian cycles in the random graph with edge probability $p = a/(n-1)$?

Proof.
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

$$E[X] = E[\sum_i X_i] = n!(a/(n-1))^n \approx (n/e)^n (a/(n-1))^n \approx (a/e)^n 1/(1-1/n)^n \approx (a/e)^n \cdot e \geq (a/e)^n$$

■

Problem 37.

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).

Proof.
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

$$\begin{aligned}
\Pr(\text{there is no p.m.}) &\leq \sum_{i=1}^{n} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

Thus for sufficiently large n, the chance that a random $n \times n$ bipartite graph does not have a perfect macthing is very small.

■

Problem 38.

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.

Proof.
The probability that a given vertex has degree exactly k is equal to $\binom{n-1}{k} p^k (1-p)^{n-1-k}$. Summing for all values of k from 0 to 10 we get the probability that a vertex has degree less than 11. If we multiply by n we get an upper bound on the probability that some vertex has degree less than 11. It suffices to prove that the latter expression (call it P) is upper bounded by a function $\varepsilon(n)$ such that $\varepsilon(n) \to 0$, as $n \to \infty$. This can be proven as follows . We use $p = (1+\varepsilon)\log n/n$, $\varepsilon > 0$, and thus $(1-p)^n \leq e^{-pn} = \frac{1}{n^{1+\varepsilon}}$, while $1/(1-p) < 2$. We also use $\binom{n-1}{k} \leq n^k$.

$$P = n\sum_{k=0}^{10}\binom{n-1}{k}p^k(1-p)^{n-1-k} \leq n(1-p)^n\sum_{k=0}^{10}n^k\frac{(1+\varepsilon)^k\log^k n}{n^k(1-p)^{k+1}} \leq n\frac{1}{n^{1+\varepsilon}}\sum_{k=0}^{10}(1+\varepsilon)^k\log^k n\, 2^{k+1} \leq \frac{11\ 2^{11}(1+\varepsilon)^{10}\log^{10} n}{n^{\varepsilon}} \to 0 \text{ when } n \to \infty.$$

where the last sum, was bounded above by 11 times its largest term. Note, that no matter how bad i overestimated, I got the desired result.

■

Problem 39.

Let $G(n,p),\ 0 \le p \le 1$, 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 $n \to \infty$,

$$\frac{cn}{\log n} \le \gamma(G),$$

where $\gamma(G)$ is the chromatic number of G.

Proof.

The probability that an n vertexgraph with edge probability $\frac{1}{2}$ has a k independent set is at most

$$\binom{n}{k} 2^{-\binom{k}{2}} \le n^k 2^{-\frac{k^2}{4}}$$

For $k = (4+\varepsilon)\lg n$, this means the chance there is an independent set larger than that is no more than

$$\begin{aligned} (n2^{-\frac{k}{4}})^k &\le (nn^{-1-\frac{\varepsilon}{4}})^k \\ &= n^{-\varepsilon k} \end{aligned}$$

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 $\gamma(G) \ge \frac{n}{5\lg n}$.

∎

Problem 40.

Show that any bipartite graph $G(X \cup Y, E)$ ($\mid X \mid = \mid Y \mid = n$) with edge probability $p = \frac{1}{2}$ has a perfect matching with high probability (i.e. with probability tending to 1 as $n \to \infty$).

Proof.
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 $\mid A \mid > \mid R(A) \mid$. We shall show that the probability that any A has $\mid R(A) \mid \leq \mid A \mid -1$ is small. For a particular A, $\mid R(A) \mid < \mid A \mid$ only if there are at least $n - \mid A \mid +1$ 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

$$\begin{aligned}
\Pr(\text{there is no p.m.}) &\leq \sum_{i=1}^{n} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} \sum_{i=n/2+1}^{n} \binom{n}{i} \binom{n}{n-i+1} (0.5)^{i(n-i+1)} \\
&< \sum_{i=1}^{n/2} n^i n^i (0.5)^{i(n-i+1)} + \sum_{i=n/2+1}^{n} n^{n-i} n^{n-i+1} (0.5)^{i(n-i+1)} \\
&= \sum_{i=1}^{n/2} (\frac{n^2}{2^{n-i+1}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^i})^{n-i+1} \\
&< \sum_{i=1}^{n/2} (\frac{n^2}{2^{n/2}})^i + \sum_{i=n/2+1}^{n} (\frac{n^2}{2^{n/2}})^{n-i+1} \\
&\to 0
\end{aligned}$$

Thus for sufficiently large n, the chance that a random $n \times n$ bipartite graph does not have a perfect matching is very small.

■

Problem 41.

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 $n \to \infty$).

Proof.
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

$$\frac{1}{n^{\varepsilon}} \to 0 \;\; as \;\; n \to \infty$$

and therefore the probability that the graph is of diameter 2 tends to 1 for large n.

∎

Problem 42.

We 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.

Proof.
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

$$\binom{n}{k} \Pr\left(\text{a particular } x_1, \ldots, x_k \text{ are not all beaten by some vertex} y\right)$$

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 $n \geq 2k^2 2^k$.

■

Problem 43.

Let $G_{n,p}$ be a undirected random graph on n vertices generated by independently including each edge with probability p. What is the expected number of (exactly) ten vertexcycles if $p = \frac{2}{n}$? How does this expectation grow with n (as n goes to ∞)?

Proof.

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

$$E \to \frac{2^{10}}{2\cdot 10} = 102.4$$

■

Problem 44.

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).

Proof.
i) This part has been solved in problem Set 6 (extra problem). 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).

■

Chapter 14

Random matrices

Problem 45.

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.

Proof.
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 $\binom{n}{\frac{n}{2}}^2$ since we can choose $\frac{n}{2}$ rows (out of a total of n) in $\binom{n}{\frac{n}{2}}$ 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

$$\binom{n}{\frac{n}{2}}^2 (1-\frac{k}{n})^{\frac{n^2}{4}}$$

We know employ Stirling's approximation formula. We get an upperbound by ignoring the square root terms.

$$\binom{n}{\frac{n}{2}}^2 (1-\frac{k}{n})^{\frac{n^2}{4}} \approx \left(\frac{(\frac{n}{e})^n}{(\frac{n}{2e})^{n/2}(\frac{n}{2e})^{n/2}}\right)^2 (1-\frac{k}{n})^{\frac{n}{k}\frac{kn}{4}} \leq 4^n e^{-\frac{kn}{4}} \tag{14.1}$$

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 equation (2) goes asymptotically to 0 when n goes to infinity. This means that the probability that a submatrix contains only 0 elements is small and therefore the probability that all submatrices contain at least one 1 is sufficiently large i.e. such a matrix exists.

■

Problem 46.

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.

Proof.
The number of possible $n \times n$ matrices is 2^{n^2}. We examine the complement of the problem i.e we would like to have, for sufficiently large n, that the number of matrices with a pair of rows (at least) that differ in less than $\frac{49n}{100}$ positions is too small (in our discussion below we would ignore constant ± 1 differences in positions of the two rows). We now estimate the number of matrices that have a pair of rows with at most $\frac{49n}{100}$ differing positions (if we wanted to be correct this should be $\frac{49n}{100} - 1$ differing positions, but as we said we ignore such constants by overestimating).

An overestimate on the number of such matrices is the following

$$\binom{n}{2} \sum_{i=0}^{\frac{49n}{100}} \binom{n}{i} 2^i \, 2^{n-i} \quad 2^{n^2-2n}$$

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 $\binom{n}{\frac{49n}{100}}$. Since $2^i \cdot 2^{n-i} = 2^n$ we can move 2^n out of the sum, which is at most twice the dominating term.

$$\frac{\binom{n}{2} 2 \cdot \binom{n}{\frac{49n}{100}} 2^n \; 2^{n^2-2n}}{2^{n^2}}$$

which is

$$\frac{\binom{n}{2} \binom{n}{\frac{49n}{100}}}{2^{n-1}}$$

For the $\binom{n}{\frac{49n}{100}}$ we use Stirling's approximation formula which finally yields (after upperbounding the square root that appears in the denominator) the following expression

$$\frac{\binom{n}{2} 2^{\alpha n}}{2^{n-1}}$$

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 $\binom{n}{\frac{49n}{100}}$ 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.

Problem 47.

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:

$$per(A) = \sum_{\text{all permutations } \sigma \text{ on } \{1,2,\ldots,n\}} \prod_{i=1}^{n} a_{i\sigma(i)}$$

Proof.

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:

$$E(\prod_{i=1}^{n} a_{i\sigma(i)}) = 1 \cdot Pr(\prod_{i=1}^{n} a_{i\sigma(i)} = 1) + 0 \cdot Pr(\prod_{i=1}^{n} a_{i\sigma(i)} = 0) = 1 \cdot 2^{-n} = 2^{-n} \tag{14.2}$$

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, useing the same observation as above. We need only to note that half the permutations are odd (and thus of sign 1) and half even (and of sign 0). (The definition of an odd or an even permutation can be found in the textbook.) Then $(-1)^{sign(\sigma)}$ is half of the times 1 and the other half -1.

$$E(det(A)) = \sum_{\sigma \text{ on } \{1,2,\ldots,n\}} (-1)^{sign(\sigma)} E(\prod_{i=1}^{n} a_{i\sigma(i)})) = \sum_{\sigma \text{ on } \{1,2,\ldots,n\}} (-1)^{sign(\sigma)} 2^{-n} = 2^{-n}\, 0 = 0$$

We got the first equality by using the formula for the expectation of the sum of random variables, and the observation that the value of the product does not depend on the permutation chosen but only on the values of the a_{ij}. The second equality is due to (1), and the third comes from the fact that $n!/2$ permutations are even and contribute each one of them an 1, while $n!/2$ are odd and contribute -1. Another way to solve this problem is expansion by minors.

Note. In a random bipartite graph (we have two sets of n vertices, the left and the right one, and a vertex of the left set is connected to a vertex of the right set with probability p, independently of the other choices), we show that the expected number of perfect matchings is $p^n n!$. The number of perfect matchings is just the permanent of a matrix with entries a_{ij} that take values 1 or 0 with probabilitites p and $1 - p$ respectively (and a value of 1 indicates an edge from vertex i of the left set to vertex j of the right one). A permutation is just an 1-1 function on $\{1, 2, \ldots, n\}$. Each permutation gives a perfect matching, and a given permutation appears with probability p^n (since an edge from vertex i of the left set to $\sigma(i)$ of the right set, for a permutation σ, appears with probability p). We have $n!$ permutations, therefore the expected number of perfect matchings is just $n!p^n$. Take $p = 1/2$ and you have another solution for the *permanent* part of the problem.

∎

Chapter 15

Randomness

Problem 48.

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) {
  m = 0;
  for(i=0 ; i < n ; i++){
    m = 2 * m + RandomBit();
  }
  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?

Proof.
(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 $\underbrace{1/2 \times 1/2 \times \ldots \times 1/2}_{n \text{ times}}$.

■

Problem 49.

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?

Proof.
(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

$$E[X] = \ln n + \gamma.$$

■

Problem 50.

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.

Proof.

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.

■

Problem 51.

Show that $n! \nmid n^n$ for $n > 2$.

Proof.
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. ■

Problem 52.

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

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

Proof.
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
```

■

Problem 53.

Consider the following algorithm.

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

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

Proof.
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 $A[1]$.

■

Problem 54.

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.

Proof.

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

Problem 55.

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.

Proof.
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%.

■

Problem 56.

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.

Proof.
(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. ■

Problem 57.

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.

Proof.
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. ■

Problem 58.

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?

Proof.
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.

■

Chapter 16

Ramsey numbers

Problem 59.

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:

$$2^k \leq R(C_3,k) \leq 3 \cdot k!$$

Proof.
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.
b) (**Exercise** : *Show that this bound can be as small as* $ek!+1$.) 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.

■

Problem 60.

For any $k > 4$, show $R(k,k) \geq 2^{k/2}$.

Proof.
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 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 $R(k,k) > 2^{k/2}$.

∎

Problem 61.

Prove 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.)

Proof.

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

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

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.

∎

Problem 62.

(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.

Proof.
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 $(n/k)n$.

■

Problem 63.

For $m,\ r,\ n$ positive integers the set $\{1,\dots,m\}$ has property $B(r,n)$ if for some collection of r subsets of size n of $\{1,\dots,m\}$ any 2-coloring of $\{1,\dots,m\}$ results in some member of the collection being monochromatic. Find a lower bound on r such that $\{1,\dots,m\}$ has property $B(r,n)$.

Proof.
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 $\{1,\dots,m\}$ 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