Chapter 7

Finite Probability Spaces

Part of chapter from
László Babai: "Discrete Mathematics" (Lecture notes, 2003, 2020, 2021, 2024)
Last updated February 16, 2024.

7.1 Notation: sets, functions, strings, closed-form expressions

Notation 7.1.1. We write $\mathbb{N} = \{1, 2, 3, \dots\}$ for the set of *natural numbers* (positive integers) and $\mathbb{N}_0 = \{0, 1, 2, \dots\} = \{0\} \cup \mathbb{N}$ for the set of non-negative integers. For $n \in \mathbb{N}_0$ we write $[n] = \{1, 2, \dots, n\}$. So $[3] = \{1, 2, 3\}$, $[1] = \{1\}$, $[0] = \emptyset$. If A and B are sets then B^A denotes the set of functions $f : A \to B$ (functions with domain A and codomain B). $|A|$ denotes the *cardinality* (size) of the set A (the number of elements of A). For instance, for $n \in \mathbb{N}_0$ we have $|[n]| = n$. A set of size k is referred to as a k-set. A *k-subset* of a set A is the set of those subsets of A that are k-sets. For a set A we write $\mathcal{P}(A) = \{B \mid B \subseteq A\}$ for the *powerset* of A (set of all subsets of A). If A is a set and $k \in \mathbb{N}_0$ then we write

$$\binom{A}{k} = \{B \subseteq A \,:\, |B| = k\} \tag{7.1}$$

(the set of k-subsets of A).

Notation 7.1.2 (Strings)**.** The *sequences* of length n of elements from a set Σ are functions $f : [n] \to \Sigma$. We can represent such a function as a *string* (or a *word*) of length n over the "alphabet" Σ. (Any finite set can be used as the alphabet.) For instance, if $\Sigma = \{2, 5, 7\}$ then $f = \overline{727752}$ is a string of length 6 over Σ. It denotes the function $f : [6] \to \Sigma$ defined by $f(1) = 7, f(2) = 2, f(3) = 7, f(4) = 7, f(5) = 5, f(6) = 2$. The overline serves to distinguish this string from the product $7 \cdot 2 \cdot 7 \cdot 7 \cdot 5 \cdot 2$. But we omit the overline when it is clear from the context that we are talking about a string rather than a product of numbers. The *empty string* is denoted by Λ; this is the unique string of length 0. For $n \in \mathbb{N}_0$, we write Σ^n to denote

the set of strings (words) of length n over the alphabet Σ. For example, if $\Sigma = \{a, X, 7\}$ then $XaX777aa \in \Sigma^8$. The set Σ^n is in a natural 1-to-1 correspondence with the set $\Sigma^{[n]}$ of functions $[n] \to \Sigma$.

We call $\mathbb{B} = \{0, 1\}$ the *Boolean alphabet*. *Boolean strings*, also called $(0, 1)$-strings, are strings over $\mathbb{B}$. So $\mathbb{B}^n$ denotes the set of $(0, 1)$-strings of length n. For instance, $\mathbb{B}^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$.
We can think of the Boolean strings of length n as representing the outcomes of the experiment where we flip n coins; the outcome of each coin flip is either "Heads" (H) or "Tails" (T). So, for instance, we interpret the string 011 as THH.

Exercise 7.1.3. (a) Prove: If A and B are finite sets then $\left|B^A\right| = |B|^{|A|}$.
(b) Let $n, k \in \mathbb{N}_0$. Let A be set of size n. We define the **binomial coefficient** $\binom{n}{k}$ by the equation

$$\binom{n}{k} = \left|\binom{A}{k}\right|. \tag{7.2}$$

Note that we did not assume $n \geq k$. It follows from the definition that if $n < k$ then $\binom{n}{k} = 0$.
(c) Prove: $\mathcal{P}(A) = 2^{|A|}$.
(d) Prove: $|\Sigma^n| = |\Sigma|^n$.

Definition 7.1.4. A **closed-form expression** is an arithmetic expression that does not involve summation ($\sum$) or product ($\prod$) symbols or ellipses (dot-dot-dots) and is made up of a "standard" set of basic functions and operations. In this course, our standard set will consist of the four arithmetic operations, taking powers, the factorial function, binomial coefficients, and the constants $0, 1, \mathrm{e}, \pi$. Complex roots of unity are also included (as powers of 1). The next exercises provide examples of non-closed-form expressions that can also be written as closed-form expressions.

Exercise 7.1.5. Let $n \in \mathbb{N}_0$. Let $S(n, 2) := \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \sum_{k=0}^{\infty} \binom{n}{2k}$. (An infinite sequence of zeros adds up to zero.) Express $S(n, 2)$ as a simple closed-form expression.
Hint. Experiment with small values of n, arrive at a conjecture, prove your conjecture.

Exercise 7.1.6. Prove: (a) $\sum_{k=0}^{\infty} \binom{n}{k} = 2^n$.
(b) (Vandermonde's identity) $\sum_{k=0}^{\infty} \binom{n}{k}^2 = \binom{2n}{n}$.
(c) Let $T(n, k) = \binom{n}{n} + \binom{n+1}{n} + \cdots + \binom{n+k}{n}$. Give a simple closed-form expression for $T(n, k)$.

Definition 7.1.7. The Fibonacci numbers $F_0, F_1, \ldots$ are defined by the **recurrence** $F_n = F_{n-1} + F_{n-2}$ (for $n \geq 2$) and the **initial values** $F_0 = 0$, $F_1 = 1$. So the sequence is $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$. For instance, $F_5 = 5$, $F_{10} = 55$, $F_{12} = 144$.

Exercise 7.1.8. Let $n, k \in \mathbb{N}_0$ and let $d = \gcd(n, k)$. Prove: $\gcd(F_n, F_k) = F_d$.

Exercise 7.1.9. Let $\phi = (1 + \sqrt{5})/2$ (the **golden ratio**) and $\overline{\phi} = (1 - \sqrt{5})/2$ (the algebraic conjugate of ϕ). Prove:

$$F_n = \frac{\phi^n - \overline{\phi}^n}{\sqrt{5}}. \tag{7.3}$$

So we have a closed-form expression for the Fibonacci numbers. This also means that we can add the Fibonacci numbers to our "standard set" without changing the set of functions that admit a closed-form expression.

Exercise 7.1.10. For $n \in \mathbb{N}_0$, let b_n denote the number of $(0,1)$-strings of length n without consecutive zeros. So for instance, the string 10011101 does not count, while the string 10111011 does count toward b_8.
Show that b_n can be expressed as a closed-form expression.

Exercise 7.1.11. For $n \in \mathbb{N}_0$, prove:

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n-k}{k} = F_{n+1}\,. \tag{7.4}$$

Exercise 7.1.12. (a) For $n \in \mathbb{N}_0$, express the sum $S(n,4) := \sum_{k=0}^{\infty} \binom{n}{4k}$ as a closed-form expression. (b) Determine, for what values of n does $S(n,4) = 2^{n-2}$ hold.

7.2 Asymptotic evaluation of sequences

In exercises like those about random graphs, one often has to estimate binomial coefficients. The following result comes in handy.

Stirling's formula.

$$n! \sim (n/\mathrm{e})^n \sqrt{2\pi n}. \tag{7.5}$$

Here the $\sim$ notation refers to *asymptotic equality:* for two sequences of numbers a_n and b_n we say that a_n and b_n are **asymptotically equal** and write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

To "evaluate a sequence a_n asymptotically" means to find a simple expression describing a function $f(n)$ such that $a_n \sim f(n)$. Stirling's formula is such an example. While such "asymptotic formulae" are excellent at predicting what happens for "large" n, they do not tell how large is large enough.

An effective (non-asymptotic) variant, giving useful results for specific values of n, is the following:

$$n! = (n/\mathrm{e})^n \sqrt{2\pi n}(1 + \theta_n/(12n)), \tag{7.6}$$

where $|\theta_n| \le 1$.

Exercise 7.2.1. Evaluate asymptotically the binomial coefficient $\binom{2n}{n}$. Show that $\binom{2n}{n} \sim c \cdot 4^n/\sqrt{n}$ where c is a constant. Determine the value of c.

We mention some important asymptotic relations from number theory. Let $\pi(x)$ denote the number of all prime numbers $\leq x$, e. g., $\pi(2) = 1$, $\pi(10) = 4$, $\pi(100) = 25$. The **Prime Number Theorem** of Hadamard and de la Vallée-Poussin (1896) asserts that

$$\pi(x) \sim x/\ln x. \tag{7.7}$$

Another important relation estimates the sum of reciprocals of prime numbers. The summation below extends over all primes $p \leq x$.

$$\sum_{p \leq x} 1/p \sim \ln \ln x. \tag{7.8}$$

In fact a stronger result holds: there exists a number B such that

$$\lim_{x \to \infty} \left(\sum_{p \leq x} 1/p - \ln \ln x \right) = B. \tag{7.9}$$

(Deduce (7.8) from (7.9).)

Exercise 7.2.2. Assuming 100-digit integers are "large enough" for the Prime Number Theorem to give a good approximation, estimate the probability that a random integer with at most 100 decimal digits is prime. (The integer is drawn with uniform probability from all positive integers in the given range.)

7.3 Finite probability space, events

Definition 7.3.1 (Probability distribution)**.** Let Ω be a non-empty finite set. A function $f : \Omega \to \mathbb{R}$ is called a **probability distribution on** Ω if it satisfies the following two conditions:

(i) $(\forall a \in \Omega)(f(a) \geq 0)$ and

(ii) $\sum_{a \in \Omega} f(a) = 1\,.$

We say that this probability distribution is **uniform** if

$$(\forall a \in \Omega) \left(f(a) = \frac{1}{|\Omega|} \right). \tag{7.10}$$

Definition 7.3.2. A **finite probability space** is a pair $\mathcal{P} = (\Omega, \Pr)$ where Ω is a non-empty finite set and $\Pr : \Omega \to \mathbb{R}$ is a probability distribution on Ω. We say that the probability space is **uniform** if $\Pr$ is the uniform distribution on Ω.

We refer to the set Ω as the **sample space** and think of it as the set of possibe outcomes of an experiment. We refer to the elements of Ω as **elementary events**.

Examples 7.3.3. 1. For $s = \overline{X_1 \dots X_n} \in \mathbb{B}^n$ let $k(s) = \sum_{i=1}^n X_i$, the number of Heads in the coin flip sequence. Let us fix a real number p in the interval $0 \le p \le 1$. Let us define the function $\Pr_p : \mathbb{B}^n \to \mathbb{R}$ by the formula

$$\Pr_p(s) = p^{k(s)} \cdot (1-p)^{n-k(s)} . \tag{7.11}$$

This equation defines a probability distribution on $\mathbb{B}^n$ (prove!).

2. Let C_{52} denote the set of 52 cards of the *standard deck*. A **poker hand** is an element of $\binom{C_{52}}{5}$, i. e., a set of 5 cards. When referring to poker hands, we always consider the uniform distribution on $\binom{C_{52}}{5}$.

Exercise 7.3.4. (a) Prove that Eq. (7.11) defines a probability distribution on $\mathbb{B}^n$. (b) For what value of p is $\Pr_p$ the uniform distribution on $\mathbb{B}^n$? (c) The number of poker hands is $\binom{52}{5}$.

Definition 7.3.5 (Events)**.** Given a finite probability space $(\Omega, \Pr)$, an **event** is a subset of Ω. We identify the elementary event $a \in \Omega$ with the event $\{a\}$.

For the event $A \subseteq \Omega$, we define the **probability** of A to be

$$\Pr(A) := \sum_{a \in A} \Pr(a) . \tag{7.12}$$

In particular, for elementary events we have $\Pr(\{a\}) = \Pr(a)$.

Exercise 7.3.6. Prove: $\Pr(\emptyset) = 0$ and $\Pr(\Omega) = 1$.

Definition 7.3.7. The **trivial events** are those with probability 0 or 1.

Exercise 7.3.8. Prove: the number of trivial events is a power of 2.

Exercise 7.3.9. In a uniform probability space, calculation of probabilities amounts to counting:

$$\Pr(A) = \frac{|A|}{|\Omega|} . \tag{7.13}$$

This is the naive notion of probability: "number of good cases divided by the number of all cases."

CONVENTION [Unspecified distribution assumed uniform] Let Ω be a non-empty finite set. If we say "pick an element at random from Ω" without specifying a probability distribution on Ω, we mean the uniform distribution, so for any $a \in \Omega$, the probability that a is being picked is $1/|\Omega|$.

Exercise 7.3.10 (Full house)**.** A poker hand is a "full house" if it consists of three cards of a kind (say three Kings) and two cards of another kind (say two 7s). We define the event "full house" as the subset of $\Omega = \binom{C_{52}}{5}$ consisting of the poker hands that are full house. Calculate the probability of the full house event. Give a simple formula involving binomial coefficients.

Exercise 7.3.11 (Coin flips)**.** Let $0 \leq p \leq 1$. Consider the probability space $(\mathbb{B}^n, \Pr_p)$ where the probability distribution $\Pr_p$ is defined by Eq. (7.11). (a) Show that (a1) $(\forall i)(\Pr(X_i = 1) = p)$ (a2) $(\forall i \neq j)(\Pr(X_i = 1 \text{ and } X_j = 1) = p^2)$ (a3) $(\forall I \subseteq [n])(\Pr((\forall i \in I)(X_i = 1)) = p^{|I|})$ (b) Determine the probabilities of the following events: (b1) $X_1 = X_2$ (b2) $X_1 \neq X_2$ (b3) $\sum_{i=1}^{n} X_i = k$. Your answers should be simple closed-form expressions of the input variables n, p, and k.

Exercise$^+$ 7.3.12. (Continued) (i) $\sum_{i=1}^{n} X_i$ is even. (Again, your answer should be a simple closed-form expression.)
(ii) For what values of p is this probability greater than $(1/2)(1 + 3^{-n})$?

Exercise 7.3.13 (Bridge)**.** In the card game of *bridge*, the standard deck of 52 cards is evenly distributed among four players called North, East, South, and West. What sample space does each of the following questions refer to: (a) What is the probability that North holds all the aces? (b) What is the probability that each player holds one of the aces? – These questions refer to uniform probability spaces. Calculate the probabilities.

The rest of this section refers to a fixed finite probability space $\mathcal{P} = (\Omega, \Pr)$.

Exercise 7.3.14 (Modular equation)**.** Let $A, B \subseteq \Omega$ be events. Prove:

$$\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B). \tag{7.14}$$

Definition 7.3.15. Events A and B are **disjoint** if $A \cap B = \emptyset$. Events A and B are **almost disjoint** if $\Pr(A \cap B) = 0$.

Exercise 7.3.16 (Union bound)**.** Let $A_1, \ldots, A_k \subseteq \Omega$ be events. Prove:

$$\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^{k} \Pr(A_i). \tag{7.15}$$

Prove also that equality holds if and only if the A_i are pairwise almost disjoint.

7.4 Conditional probability, probability of causes

7.4.1 Conditional probability

Definition 7.4.1 (Conditional probability)**.** If A and B are events and $\Pr(B) > 0$ then the **"probability of A, given B,"** denoted $\Pr(A \mid B)$, is given by the equation

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}\,. \tag{7.16}$$

(B is the *condition*.)

In all exercises involving conditional probabilities, we assume that the condition has positive probability.

Exercise 7.4.2. Let $B \subseteq \Omega$ with $\Pr(B) > 0$. For $a \in \Omega$, let $\Pr'(a) = \Pr(a \mid B)$.
(a) Prove: $(\Omega, \Pr')$ is probability space. Moreover, for all $A \subseteq \Omega$ we have $\Pr'(A) = \Pr(A \mid B)$.
(b) Prove: $(B, \Pr'_{|B})$ is probability space. Moreover, for all $A \subseteq B$ we have $\Pr'(A) = \Pr(A \mid B)$.
(Here, $\Pr'_{|B}$) is the restriction of the function $\Pr'$ to B.)

Note that

$$\Pr(A \cap B) = \Pr(A \mid B) \Pr(B)\,. \tag{7.17}$$

Exercise 7.4.3 (Bayes's Theorem)**.**

$$\Pr(B \mid A) = \frac{\Pr(A \mid B) \cdot \Pr(B)}{\Pr(A)}\,. \tag{7.18}$$

The theorem is named after Reverend Thomas Bayes (1702–1761). $P(B)$ is called the *prior probability* (the probability of event B before knowing that event A has occurred), and $P(B \mid A)$ the *posterior probability* of B given A. $P(A)$ is called the *marginal probability*.

Exercise 7.4.4. Prove: $\Pr(A \cap B \cap C) = \Pr(A \mid B \cap C) \Pr(B \mid C) \Pr(C)$.

Exercise 7.4.5. We roll three dice. Each die shows a number from 1 to 6. (a) What is the probability that the first die shows 5? (b) What is the probability that the sum of the three numbers shown is 9? (c) What is the probability that the first die shows 5 given that the sum of the three numbers shown is 9? (d) What is the probability space in this problem? How large is the sample space? (In accordance with our convention, the distribution under consideration is uniform.)

Definition 7.4.6. A **partition** of Ω is a set of pairwise disjoint events $H_1, \ldots, H_k$ of positive probability, covering Ω:

$$\Omega = H_1 \cup \ldots \cup H_k, \qquad H_i \cap H_j = \emptyset, \qquad (\forall i)(\Pr(H_i) \neq 0)\,. \tag{7.19}$$

The sets H_i are the *blocks* (or *parts*) of the partition.

In computer technology, blocks of a partition have also come to be called "partitions." In this course, please avoid this confusing terminology.

Exercise 7.4.7 (Theorem of Complete Probability)**.** Let $(H_1, \ldots, H_k)$ be a partition of Ω and let $A \subseteq \Omega$ be an event. Then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A \mid H_i) \Pr(H_i). \tag{7.20}$$

The significance of this formula is that the conditional probabilities are sometimes easier to calculate than the left-hand side.

7.4.2 Probability of causes

We illustrate this concept on some examples.

Consider a test for disease D. A *false positive* is a positive outcome of the test while the patient does not have disease D, and a *false negative* is a negative outcome while the patient does have the disease. Note that the probability of a false positive is a conditional probability (probability of positive outcome given that the disease is not present) and analogously, the probability of a false negative is also a conditional probability.

Exercise 7.4.8. A random member x of a population W is tested for disease D. We know that the test has 1% false positives and 1% false negatives when applied to random members of the population W. We also know that 1% of the population W suffers from disease D. Person x tests positive. What is the probility that x has disease D?

With such a highly reliable test, one would think the probability of being sick given a positive test result is high. In fact, it turns out that the probability is 50%. (Work it out.)

This striking exercise is a (significant) modification of an example stated by Chris Wiggins in an article in the *Scientific American* (Dec 4, 2006), with the provocative title *"What is Bayes's theorem, and how can it be used to assign probabilities to questions such as the existence of God? What scientific value does it have?"*

Wiggins begins his article with the following question.

> A patient goes to see a doctor. The doctor performs a test with 99 percent reliability–that is, 99 percent of people who are sick test positive and 99 percent of the healthy people test negative. The doctor knows that only 1 percent of the people in the country are sick. Now the question is: if the patient tests positive, what are the chances the patient is sick?

At first glance this question seems identical with Ex. 7.4.8. Indeed, professor Wiggins informs the reader of the counterintuitive answer that the chance in question is 50%.

In fact, the chance is likely to be considerably higher. The data point that "only 1 percent of the people in the country are sick" has limited relevance; the relevant data would be the prevalence of disease D among "patients," i. e., individuals who "go to see a doctor," presumably because of some symptoms. Perhaps we should further restrict the population to those patients with symptoms compatible with disease D, or even indicative of the possibility of disease D – why else would the doctor run a test for that particular disease? Presumably the prevalence of D is considerably higher among this restricted population than among "the people in the country."

Exercise 7.4.9. In Ex. 7.4.8, what is the probability that x has disease D if the test comes back negative? State your answer as a fraction reduced to its smallest terms (i. e., the numerator and the denominator are relatively prime).

Exercise 7.4.10. In Ex. 7.4.8, what is the probability that comes back (a) positive (b) negative, assuming the prevalence of disease D among population W is 3% ?

Exercise 7.4.11. In Ex. 7.4.8, the doctor opts for a less expensive test. The test still has 1% chance of a false negative, but the chance of a false positive is 3%. How does this affect the probability that x has disease D if the test comes back (a) positive (b) negative?

Exercise 7.4.12. Diseases A and B have similar symptoms. Let W be the population of all patients showing these symptoms. The two diseases can only be differentiated by costly tests. We know (from sampling the population and performing these costly tests) that 70% of W have disease A, 25% have disease B, and 5% have some other disease. We consider the effectiveness of treatment T. We know that 60% of the patients with disease A respond to T, while only 12% of the patients with disease B respond to treatment T. From the rest of the population W, 40% respond to treatment T. Answer the following questions. State the exact value of each required probability as a fraction reduced to it lowest terms.

(a) A new patient arrives at the doctor's office. The doctor determines that the patient belongs to W. What is the probability that the patient will respond to treatment T?

(b) The patient's insurance will not pay for the expensive tests to differentiate between the possible causes of the symptoms. The doctor bets on treatment T. A week later it is found that the patient did respond to the treatment. What is the probability that the patient had disease A? **Show all the intermediate results you need to compute.**

(c) What is the probability space to which the discussion above refers?

Warning: Your answer to (c) needs to be simple. You cannot base it on calculations you performed in (a) and (b); those calculations don't make sense without having previously defined a probability space. So answering (c) has to precede answering (a) and (b).

7.5 Independence, positive and negative correlation of a pair of events

Definition 7.5.1. Events A and B are **independent** if $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.

The intuitive meaning of this definition is supported by the following observation.

Exercise 7.5.2. Assume $\Pr(B) > 0$. Then A and B are independent $\iff \Pr(A \mid B) = \Pr(A)$.

Definition 7.5.3. The **complement** of the event A is the event $\overline{A} = \Omega \setminus A$.

Exercise 7.5.4 (Independence of complement)**.** If A and B are independent events then $\overline{A}$ and B are also independent.

Exercise 7.5.5 (Independence of a trivial event)**.** Let A be any event and B a trivial event. Show that A and B are independent.

Exercise 7.5.6. If we roll a die, are the following events independent: "the number shown is odd"; "the number shown is a square"?

The following result is at the heart of the proof of the *Fundamental Theorem of Arithmetic* (positive integers have unique prime factorization), first proved in Euclid's *Elements* about 2300 years ago.

Theorem 7.5.7 (Euclid's Lemma)**.** *If a prime number p divides a product ab where a and b are integers then p divides a or p divides b.*

Exercise 7.5.8. Let us consider a uniform probability space over a sample space whose cardinality is a prime number. Prove that no two non-trivial events can be independent. Explicitly use Euclid's Lemma.

Exercise 7.5.9. Assume there exist two nontrivial independent events in our probability space. Prove: $|\Omega| \geq 4$.

Definition 7.5.10 (Positively/negatively correlated events)**.** The events A and B are said to be **positively correlated** if $\Pr(A \cap B) > \Pr(A)\Pr(B)$. They are **negatively correlated** if $\Pr(A \cap B) < \Pr(A)\Pr(B)$.

Exercise 7.5.11. Let $A, B \subseteq \Omega$. Assume $\Pr(B) > 0$. Then A and B are positively (negatively) correlated if and only if $\Pr(A \mid B) > \Pr(A)$ ($\Pr(A \mid B) < \Pr(A)$, resp.).

Exercise 7.5.12. Pick a number x at random from $[n] = \{1, \ldots, n\}$. Consider the following two evens: A_n is the event that x is even; B_n is the event that x is divisible by 3. Determine whether A_n and B_n are positively correlated, independent, or negatively correlated. Your answer should be a function of (n mod 6) (you need to list 6 cases). – What is the probability space for this experiment?

Exercise 7.5.13. Let $A \subseteq B \subseteq \Omega$. If A and B are independent then one of them is trivial.

Exercise 7.5.14. Let A, B be events. Suppose $A \cap B$ and $A \cup B$ are independent. Does it follow that A or B must be trivial? If true, give a short proof. If false, minimize the size of the sample space of your counterexample.

Exercise 7.5.15. Let A be an event. Prove: A and A are independent if and only if A is a trivial event. If A is nontrivial then A and A are positively correlated.

7.6 Independence of multiple events

Definition 7.6.1. Events $A_1, \ldots, A_k$ are **pairwise independent** if $(\forall i \neq j)(A_i$ and A_j are independent).

Definition 7.6.2 (Independence of 3 events)**.** Events $A, B, C \subseteq \Omega$ are **(fully) independent** if

(i) A, B, C are pairwise independent

(ii) $\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$.

If we say "A, B, C are independent," it means they are fully independent. The term "fully" can be added to emphasize that we are not talking about pairwise independence. Another term for independence of A, B, C is that they are **mutually independent**.

Exercise 7.6.3 (Independence of complement)**.** If A, B, and C are independent events then $\overline{A}, B$ and C are also independent.

Exercise 7.6.4 (Independence of trivial event)**.** If A, B are independent and C is a trivial event then A, B, C are indepenedent.

Exercise 7.6.5 (Independence of intersection, union)**.** If A, B, C are independent events then the following pairs of events are also independent: (a) $A \cap B, C$ (b) $A \cup B, C$.

Exercise 7.6.6. Assume there exist three nontrivial independent events in our probability space. Prove: $|\Omega| \geq 8$.

Exercise 7.6.7. (a) Show that if three events are pairwise but not fully independent then none of them is trivial. (b) Define a probability space and three events in it that are pairwise but not fully independent. Compute the relevant probabilities. Make your sample space as small as possible.

Definition 7.6.8. A **balanced event** is an event of probability 1/2.

Exercise 7.6.9. (a) Define a probability space and three events, A, B, C, in it that satisfy $\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$ but are not independent. Compute the relevant probabilities. Make your sample space as small as possible. (b) Same as (a) with the additional requirement that the events be balanced.

We now wish to define, for a list of k events, what it means to be independent. Definition 7.6.2 suggests the following inductive definition. A **k-subset** is a subset of size k.

Definition 7.6.10. We say that the events $A_1, \ldots, A_n$ are **k-wise independent** if for all k-subsets $I \subseteq [n]$, the list $(A_i \mid i \in I)$ of events is independent.

Definition 7.6.11 (Independence: inductive definition)**.** Let $k \geq 3$. We say that the events $A_1, \ldots, A_k$ are independent if

(i) $A_1, \ldots, A_k$ are $(k-1)$-wise independent

(ii) $\Pr\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} \Pr(A_i)$

This definition is inductive: once we know what it means for $k-1$ events to be independent, the definition tells us what it means for k events to be independent. The base case is $k = 2$. But we could even take go further: we declare that any event alone ($k = 1$) is indepoendent; this would be the base case, and then Def. 7.6.11 will take effect for $k \geq 2$.

Next we give an alternative, non-inductive definition.

Definition 7.6.12 (Independence: explicit definition)**.** Events $A_1, \ldots, A_k$ are **independent** if for all subsets $I \subseteq [k]$ we have

$$\Pr\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \Pr(A_i)\,. \tag{7.21}$$

For this definition to make sense, we need to understand what $\bigcap_{i\in I} A_i$ means when $I = \emptyset$. If we intersect more sets, we get a smaller set, so it is natural to define the intersection of an empty list of sets to be as large as possible. So this definition needs to refer to a "largest set" which we call the "universe"; all sets we consider are subsets of the universe. In the context of events in a probability space $(\Omega, \Pr)$, this universe is, naturally, the sample space Ω.

Let us now review the definition of intersection of a list of sets.

Definition 7.6.13 (Intersection of a list of sets)**.** Let us fix a set Ω, to be referred to as the "universe." Let I be a (possibly empty) set and let $(A_i \mid i \in I)$ be a list of subsets of the universe. Then

$$\bigcap_{i\in I} A_i = \{x \in \Omega \mid (\forall i \in I)(x \in A_i)\} \tag{7.22}$$

Exercise 7.6.14. Based on Def. 7.6.13, verify the following.

(a) $\bigcap_{i\in\emptyset} A_i = \Omega$

(b) If $I = \{j\}$ (so $|I| = 1$) then $\bigcap_{i\in I} A_i = A_j$.

Exercise 7.6.15 (Equivalence of definitions of independence)**.** Show that our two definitions of independence of k events, Def 7.6.11 and Def 7.6.12, are equivalent.

How many conditions do we need to verify to establish that a list of k events is independent? Definition 7.6.12 tells us that we need to verify a condition for each subset $I \subseteq [k]$. This means 2^k conditions. In fact, a bit fewer will suffice: $k + 1$ of these are automatically satisfied, as stated in the following exercise.

Exercise 7.6.16. Show that for $|I| \leq 1$, Eq. (7.21) always holds, so in verifying the independence of a list of events, we only need to verify Eq. (7.21) for $|I| \geq 2$. This means verifying $2^k - k - 1$ conditions.

Again, for emphasis, independent events are also called *fully* independent, or *mutually* independent.

Exercise 7.6.17 (Independence of complements-1)**.** Prove: if $A_1, A_2, \ldots, A_k$ are independent events then $\overline{A}_1, A_2, \ldots, A_k$ are independent events.

Exercise 7.6.18 (Independence of complements-2)**.** For an event A define $A^1 = A$ and $A^{-1} = \overline{A}$. Prove: if $A_1, \ldots, A_k$ are independent events and $\epsilon_1, \ldots, \epsilon_k \in \{1, -1\}$ then $A_1^{\epsilon_1}, \ldots, A_k^{\epsilon_k}$ are independent events.

Exercise 7.6.19. Assume there exist k nontrivial independent events in our probability space. Prove: $|\Omega| \geq 2^k$.

Exercise 7.6.20. For all $n \geq 2$, construct a probability space and n balanced events (events of probability $1/2$) such that the events are not independent but they are $(n-1)$-wise independent. Minimize the size of the sample space.

The size of the sample space is a resource in computer science, which we wish to minimize. If we need pairwise independence only, rather than full independence, we can do much better than the 2^k size of the sample space. By the **size of a probability space** we mean the size of its sample space.

Exercise* 7.6.21 (Small sample space for pairwise independent events-1)**.**

(a) Let $k \geq 3$. Construct a probability space of size $k + 1$ and k pairwise independent nontrivial events in that space.

(b) Let $k \geq 1$. Construct a probability space of size $\leq 2k$ and k pairwise independent *balanced* events (events of probability $1/2$) in that space.

The following exercise may help solve Ex. 7.6.21 (b).

Exercise* 7.6.22 (Small sample space for pairwise independent events-2)**.** Let $\ell \geq 1$.

(i) For $k = 2^\ell - 1$, construct a uniform probability space of size $k + 1$ with k pairwise independent balanced events.

(ii) Same for k a prime number of the form $k = 4t - 1$.

Exercise** 7.6.23 (Lower bound for pairwise independent events)**.** Assume there exist k pairwise independent nontrivial events in our probability space. Prove: $|\Omega| \geq k + 1$. Note: Ex. 7.6.21 (a) shows that this bound is tight.

Exercise* 7.6.24. Let $1 \leq k \leq n - 1$.

(a) Construct a sample space Ω and n events that are k-wise independent but no $k + 1$ of the events are independent.

(b) Solve item (a) under the additional constraint that each of the n events be balanced.

(Hint for part (a). Take a k-dimensional vector space W over a finite field of order $q \geq n$. Select n vectors from W so that any k are linearly independent. Let W be the sample space.)

Exercise 7.6.25 (Independence of Boolean combinations of groups of events)**.** Prove: if the five events A, B, C, D, E are independent then the three events $A \setminus B$, $C \cup D$, and E are independent as well. Formulate a general statement, for n events grouped into blocks.

Exercise 7.6.26 (A trick problem)**.** We have n balls colored red, blue, and green (each ball has exactly one color and each color occurs at least once). We select k of the balls with replacement (independently, with uniform distribution). Let A denote the event that the k balls selected have the same color. Let p_r denote the conditional probability that the first ball selected is red, assuming condition A. Define p_b and p_g analogously for blue and green outcomes. Assume $p_r + p_b = p_g$. Prove: $k \leq 2$. Show that $k = 2$ is actually possible.

7.7 Constructing independent events: Product spaces

For $i \in [k]$, let $(\Omega_i, \Pr_i)$ be finite probability spaces. Let $\Omega = \Omega_1 \times \cdots \times \Omega_k$. To $a = (a_1, \ldots, a_k) \in \Omega$, let us assign the value

$$\Pr(a) = \prod_{i=1}^{k} \Pr_i(a_i)\,. \tag{7.23}$$

Exercise 7.7.1. Prove: this assigment defines a probability space $(\Omega, \Pr)$.

The probablity space $(\Omega, \Pr)$ is called the **product** of the spaces $(\Omega_i, \Pr_i)$.

Let $\pi_i : \Omega \to \Omega_i$ denote the i-th projection: $\pi_i(a_1, \ldots, a_k) = a_i$.

Let $A_i \subseteq \Omega_i$. The **lifting** of A_i is the set

$$\widetilde{A}_i := \pi_i^{-1}(A_i) = \{(a_1, \ldots, a_k) \mid (\forall j)(a_j \in \Omega_j), a_i \in A_i\}\,.$$

Exercise 7.7.2. $\Pr(\widetilde{A}_i) = \Pr_i(A_i)$ and the events $\widetilde{A}_i$ $(i \in [k])$ are independent.

We now have the following corollary.

Let $\Omega = \{0,1\}^k$ be the set of $(0,1)$-strings of length k. Denote the elements of Ω by $(x_1, \ldots, x_k)$ where $x_i \in \{0,1\}$.

Exercise 7.7.3 (Coin flips with variable biases)**.** (a) Prove: Given the real numbers $0 \leq p_i \leq 1$ for $i \in [k]$, there exists a probability distribution Pr on $\Omega = \{0,1\}^k$ such that for $i \in [k]$ we have $\Pr(x_i = 1) = p_i$ and the k events "$x_i = 1$" are independent. (b) Prove that this distribution is unique.

7.8 Random graphs: The Erdős–Rényi model

Let Γ_n denote the set of graphs with vertex set $V = [n]$.

Exercise 7.8.1. $|\Gamma_n| = 2^{\binom{n}{2}}$

For $e = \{i, j\} \in \binom{V}{2}$ let p_e be a real number, $0 \le p_e \le 1$.

Let $A_e = \{\mathcal{G} \in \Gamma_n \mid e \in E(\mathcal{G})\}$. This is the event that $i \sim_{\mathcal{G}} j$.

Let $\mathbf{G}(n, (p_e \mid e \in \binom{V}{2}))$ denote the probability distribution on Γ_n such that

$$(\forall e \in \binom{V}{2})(\Pr(A_e) = p_e) \tag{7.24}$$

and these $\binom{n}{2}$ events are independent. Such a probability distribution exists (and is unique) according to Ex. 7.7.3.

Let $n \in \mathbb{N}$ and $0 \le p \le 1$. The **Erdős–Rényi model** $\mathbf{G}(n, p)$ is defined as the special case of the $\mathbf{G}(n, (p_e \mid e \in \binom{V}{2}))$ model where $(\forall e \in \binom{V}{2})(p_e) = p$. In other words, in the $\mathbf{G}(n, p)$ model we have a fixed set V of n vertices and for every pair i, j of distinct vertices, $\Pr(i \sim j) = p$ and these $\binom{n}{2}$ events are independent.

In the next sequence of problems we consider the uniform probability space over the sample space Γ_n. This is called the **uniform Erdős–Rényi model** of "random graphs."

Exercise 7.8.2 (Random graphs)**.** Let $A(i, j)$ denote the event that vertices i and j are adjacent ($1 \le i, j \le n$, $i \ne j$). Note that $A(i, j) = A(j, i)$ so we are talking about $\binom{n}{2}$ events.

(a) Determine $\Pr(A(i, j))$.

(b) Prove that these $\binom{n}{2}$ events are independent.

(c) What is the probability that the degrees of vertex 1 and vertex 2 are equal? Give a simple closed-form expression.

(d) If p_n denotes the probability calculated in item (c), prove that $p_n\sqrt{n}$ tends to a finite positive limit and determine its value.

(e) How are the following two events correlated: $A_n =$"vertex 1 has degree 3"; $B_n =$"vertex 2 has degree 3"? Find the limit of the ratio $\Pr(A_n \mid B_n)/\Pr(A_n)$ as $n \to \infty$.

Definition 7.8.3. Let $G = (V, E)$ be a graph. The **distance** $\mathrm{dist}(u, v)$ of vertices u and v is the length of a shortest path between them. The **diameter** of G is the maximum distance:

$$\mathrm{diam}(G) = \max_{u,v \in V} \mathrm{dist}(u, v)\,. \tag{7.25}$$

If G is disconnected, we say that $\mathrm{diam}(G) = \infty$.

Exercise 7.8.4. Prove: almost all graphs have diameter 2.

Explanation. Let p_n denote the probability that a random graph on n vertices has a certain property. We say that **almost all graphs** have the property if $\lim_{n\to\infty} p_n = 1$.

7.9 Random variables, expected value, indicator variables, Bernoulli trials

Definition 7.9.1. A **random variable** is a function $X : \Omega \to \mathbb{R}$.

We say that X is **almost constant** if for some $u \in \mathbb{R}$, $\Pr(X = u) = 1$.

Definition 7.9.2. The **expected value** of a random variable X is

$$\mathrm{E}(X) = \sum_{a \in \Omega} X(a) \Pr(a) . \tag{7.26}$$

For $u \in \mathbb{R}$, we shall refer to the event "$X = u$," meaning the set $\{a \mid X(a) = u\}$. So the expression $\Pr(X = u)$ refers to the probability of this set.

Exercise 7.9.3 (Alternative definition of the expected value)**.** Prove:

$$\mathrm{E}(X) = \sum_{u \in \mathbb{R}} u \cdot \Pr(X = u) = \sum_{u \in \mathrm{range}(X)} u \cdot \Pr(X = u) . \tag{7.27}$$

Here $\mathrm{range}(X)$ denotes the range of X:

$$\mathrm{range}(X) = \{X(a) \mid a \in \Omega\} . \tag{7.28}$$

Exercise 7.9.4. The middle term in Eq. 7.27 seems like an infinite sum. Verify that all but a finite number of terms are zero.

Exercise 7.9.5.

$$\min X \leq \mathrm{E}(X) \leq \max X. \tag{7.29}$$

Throughout these notes, X, Y, Z, ϑ, and their subscripted versions refer to random variables.

Let us fix a probability space $\mathcal{P}(\Omega, \Pr)$. All random variables below refer to this probability space, unless the space is specified more concretely. Note that random variables over $\mathcal{P}$ can be added and can be multiplies by real numbers (scalars); they form a real vector space.

Exercise 7.9.6 (Additivity of expectation)**.** Let $X_1, \ldots, X_k$ be random variables. Then

$$\mathrm{E}(X_1 + \cdots + X_k) = \sum_{i=1}^{k} \mathrm{E}(X_i) \tag{7.30}$$

Proof: $\mathrm{E}\left(\sum_{i=1}^{k} X_i\right) = \sum_{a \in \Omega} (X_1(a) + \cdots + X_k(a)) \Pr(a) = \sum_{i=1}^{k} \sum_{a \in \Omega} X_i(a) \Pr(a) = \sum_{i=1}^{k} \mathrm{E}(X_i).$ □

Exercise 7.9.7 (Linearity expectation)**.** If $c_1, \ldots, c_k$ are constants (real numbers) then

$$\mathrm{E}\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} c_i \mathrm{E}(X_i). \tag{7.31}$$

Definition 7.9.8. An **indicator variable** is a (0,1)-valued random variable (its values are 0 or 1). Indicator variables are also called **Bernoulli trials**; an outcome of 1 is considered "success" and 0 "failure."

Definition 7.9.9. The **indicator of an event** $A \subseteq \Omega$, also called the **characteristic function** of the event, is the function $\vartheta_A : \Omega \to \{0, 1\}$ given by

$$\vartheta_A(a) = \begin{cases} 1 & \text{for } a \in A \\ 0 & \text{for } a \notin A \end{cases}$$

So the indicator of an event A is a **Bernoulli trial** with A being the event of "success." In particular, the probability of success is $\Pr(A)$.

Exercise 7.9.10 (Bijection between events and indicator variables)**.** If T is an indicator variable then there is a unique event A such that $T = \vartheta_A$.

Exercise 7.9.11. The expected value of an indicator variable is the probability of the event it indicates:

$$\mathrm{E}(\vartheta_A) = \Pr(A)\,. \tag{7.32}$$

Indicator variables are particularly useful if we want to count events, as demonstrated by several of the exercises at the end of this section.

Exercise 7.9.12. (a) Every random variable X is a linear combination of indicator variables. (b) Given a random variable X there exist functions $f_1, \ldots, f_k$ such that the random variables $X_i := f_i(X)$ are indicator variables and X is a linear combination of the X_i.

Exercise 7.9.13. Let $Y = \sum_{i=1}^{n} X_i$ where X_i is a Bernoulli trial with probability p_i of success. Then $\mathrm{E}(Y) = \sum_{i=1}^{n} p_i$.

We say that X is **nonnegative** if $X(a) \geq 0$ for all $a \in \Omega$.

Theorem 7.9.14 (Markov's Inequality)**.** *If X is nonnegative then $\forall a > 0$,*

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

Proof: Let $m = \mathrm{E}(X) > 0$. Then $m = \sum_{u \in \mathbb{R}} u \cdot \Pr(X = u) \geq$
$\geq \sum_{u \geq a} u \cdot \Pr(X = u)$ (we just omitted some terms; all terms are nonegative)
$\geq a \cdot \sum_{u \geq a} \Pr(X = u) = a \cdot \Pr(X \geq a)$ (sum of disjoint events).
So we have $m \geq a \Pr(X \geq a)$. □

Exercise 7.9.15 (Poker hand)**.** (a) What is the expected number of Aces in a poker hand? (b) What is the expected number of Spades?

Exercise 7.9.16 (Flipping coins)**.** We flip a biased coin n times. The coin comes up Heads with probability p and Tails with probability $1 - p$. What is the expected number of runs of k heads in a string of n coin-flips? (A "run of k heads" means a string of k consecutive heads. Example: the string HHTHTTHHHT has 3 runs of 2 heads.) Prove your answer! *Hint.* Indicator variables.

Exercise 7.9.17 (Lottery)**.** Suppose in a lottery you have to pick five different numbers from 1 to 90. Then five winning numbers are drawn. If you picked two of them, you win 20 dollars. For three, you win 150 dollars. For four, you win 5,000 dollars, and if all the five match, you win a million. (a) What is the probability that you picked exactly three of the winning numbers? (b) What is your expected win? (c) What does Markov's inequality predict about the probability that you'll win at least 20 dollars? (d) What is the actual probability that this happens?

Exercise 7.9.18 (Club of 2000)**.** As a matter of long-standing tradition, the Moonwatchers' Club of Onyx, NA, serves vodka legally to all of its members. Throughout the year 2020, the club had 2000 members. One of the club members wrote the following in their diary on June 27, 2020. "The managment of the club just announced that two weeks from now they will distribute membership cards numbered 1 through 2000 to the members at random. Members whose card number happens to coincide with their year of birth receive valuable gifts. How exciting!" Determine the expected number of lucky members just before the managment shuffles the cards. State the role of the vodka in your calculation. State the size of the sample space for this experiment.

Exercise 7.9.19 (Random graphs)**.** Consider a random graph G with n vertices.

(a) What is the expected number of edges in G?

(b) What is the expected number of triangles?

(c) What is the expected number of cycles of length k?

(d) Show that the expected number of Hamilton cycles (cycles of length n) is large; it is greater than 100^n for all sufficiently large n.

Exercise 7.9.20 (Distinct prime divisors)**.** Let n be a random integer, chosen uniformly between 1 and N. What is the expected number of distinct prime divisors of n? Show that the result is asymptotically equal to $\ln \ln N$ (as $N \to \infty$).

Exercise 7.9.21 (Mismatched letters)**.** The boss writes n different letters to n addressees whose addresses appear on n envelopes. The careless secretary puts the letters in the envelopes at random (one letter per envelope). Determine the expected number of those letters which get in the right envelope. State the size of the sample space for this problem.

Exercise 7.9.22 (Marbles in cups)**.** Kiara has n cups and n marbles. She puts each marble in a randomly selected cup, regardless of whether the cup already has marbles in it. What is the expected number of cups left empty? (a) Give a simple expression in terms of n. (b) Asymptotically evaluate your answer. (Find a very simple expression that is asymptotically equal to your answer.)

Exercise 7.9.23 (Counting cycles in permutations)**.** For a permutation π of the set $[n]$, let $c_k(\pi)$ denote the number of k-cycles in the cycle decomposition of π. (For instance, if $n = 7$ and $\pi = (18)(256)(3)(47)(9)$ then $c_1(\pi) = 2$, $c_2(\pi) = 2$, $c_3(\pi) = 1$, and $c_k(\pi) = 0$ for all $k \neq 1, 2, 3$.) Pick π at random from all permutations of $[n]$. (a) Calculate $E(c_k(\pi))$. Your answer should be a very simple expression (no factorials, no binomial coefficients, no summation). (b) Calculate the expected number of cycles (including cycles of length 1) in the cycle decomposition of a random permutation. (This will be a simple sum, not a closed-from expression.) Prove that this number is $\sim \ln n$.

7.10 Variance, covariance, Chebyshev's Inequality

Definition 7.10.1. The k^{th} **moment** of X is $\mathrm{E}(X^k)$. The k^{th} **central moment** of X is the k^{th} moment of $X - \mathrm{E}(X)$, i. e., $\mathrm{E}((X - \mathrm{E}(X))^k)$.

Definition 7.10.2. The **variance** of X is its second central moment, $\mathrm{Var}(X) := \mathrm{E}((X - \mathrm{E}(X))^2)$.

Note that the variance is always nonnegative.

Exercise 7.10.3. Prove: $\mathrm{Var}(X) \geq 0$ and $\mathrm{Var}(X) = 0$ if and only if X is almost constant.

Definition 7.10.4. The **standard deviation** of X is $\sigma(X) := \sqrt{\mathrm{Var}(X)}$.

Exercise 7.10.5. (Invariance under shifts.) Prove that if c is a constant then $\mathrm{Var}(X) = \mathrm{Var}(X + c)$; and consequently, $\sigma(X) = \sigma(X + c)$.

Exercise 7.10.6. Prove: if c is a constant then $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$; and consequently, $\sigma(cX) = |c|\sigma(X)$.

Exercise 7.10.7. $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

Corollary 7.10.8 (Cauchy–Schwarz inequality)**.** $(\mathrm{E}(X))^2 \leq \mathrm{E}(X^2)$. □

Proof of Observation: Let $m = \mathrm{E}(X)$. Then $\mathrm{Var}(X) = \mathrm{E}((X-m)^2) = \mathrm{E}(X^2 - 2Xm + m^2) = \mathrm{E}(X^2) - 2m\mathrm{E}(X) + \mathrm{E}(m^2) = \mathrm{E}(X^2) - 2mm + m^2 = \mathrm{E}(X^2) - m^2$. □

Chebyshev's inequality tells us that random variables don't like to stray from their expected value by more than a small multiple of their standard deviation.

Theorem 7.10.9 (Chebyshev's Inequality)**.** *Let $m = \mathrm{E}(X)$. Then for any number $a > 0$,*

$$\Pr(|X - m| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}. \tag{7.33}$$

Proof: Let $Y = (X - m)^2$. Then, by definition, $\mathrm{E}(Y) = \mathrm{Var}(X)$. We apply Markov's Inequality to the nonnegative random variable Y: $\Pr(|X-m| \geq a) = \Pr(Y \geq a^2) \leq \mathrm{E}(Y)/a^2 = \mathrm{Var}(X)/a^2$. □

Exercise 7.10.10. In its more common form the Cauchy–Schwarz inequality asserts that for any real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$ we have

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \geq \left(\sum_{i=1}^{n} x_i y_i\right)^2. \tag{7.34}$$

Deduce this inequality from Corollary 7.10.8.

Exercise 7.10.11. Prove: if the k^{th} moment of X is zero for all odd integers $k > 0$ then $\Pr(X = u) = \Pr(X = -u)$ for all $u \in \mathbb{R}$.

Definition 7.10.12 (Covariance)**.** The **covariance** of the random variables X, Y is defined as

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X) \cdot \mathrm{E}(Y)\,. \tag{7.35}$$

Exercise 7.10.13. $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$

Definition 7.10.14. We say that X and Y are **positively correlated** if their covariance is positive; they are **uncorrelated** if their covariance is zero; and **negatively correlated** if their covariance is negative.

Exercise 7.10.15 (Aces vs. Spades)**.** (a) Consider a poker hand. Let X denote the number of Aces and Y the number of Spades in the hand. Show that X and Y are uncorrelated. Show all your work. The best way to solve this problem is by solving part (b) and avoiding all numerical calculation.

(b) Generalize the problem to a deck of rs cards where there are r cards of each kind (e. g., r Aces) and and s cards in a suit (e. g., s Spades). A generalized poker hand will have k cards ($1 \leq k \leq rs$).

We define independence of random variables in the next section but we warn in advance that for a pair of random variables, independence is a stronger condition than being uncorrelated.

Exercise 7.10.16 (Events vs. indicator variables)**.** The events A, B are positively correlated if and only if the corresponding indicator variables ϑ_A and ϑ_B are positively correlated; A and B are independent if and only if ϑ_A and ϑ_B are uncorrelated; and A and B are negatively correlated if and only if ϑ_A and ϑ_B are negatively correlated.

We often deal with sums of random variables. Next we give a formula for the variance of such a sum.

Exercise 7.10.17 (Variance of sum)**.** Let $Y = X_1 + \cdots + X_n$ be a sum of random variables. Then

$$\mathrm{Var}(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{i\neq j}\mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\cdot\sum_{1\leq i<j\leq n}\mathrm{Cov}(X_i, X_j)\,. \tag{7.36}$$

Corollary 7.10.18 (Additivity of variance)**.** *If $X_1, \ldots, X_n$ are pairwise uncorrelated random variables then*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathrm{Var}(X_i)\,. \tag{7.37}$$

In particular, this equation holds if the variables are pairwise independent (see Ex. 7.11.3).

Exercise 7.10.19 (Variance of the number of triangles)**.** Let X_n denote the number of triangles in a random graph with n vertices.

(a) Determine $\mathrm{E}(X_n)$.

(b) Determine $\mathrm{Var}(X_n)$. Your answer should be a closed-form expression in terms of n.

(c) Asymptotically evaluate your answer to (b). Your answer should be of the form $\mathrm{Var}(X_n) \sim an^b$. Determine the constants a and b. *Hint:* Write X_n as a sum of indicator variables.

Exercise 7.10.20 (Limit on strongly negatively correlated events)**.** (a) Suppose the events $A_1, \ldots, A_m$ are balanced (have probability 1/2) and for each $i \neq j$, $\Pr(|A_i \cap A_j| \leq 1/5$. Prove: $m \leq 6$. (b) Generalize the statement to events of probability p, with $p^2 - \epsilon$ in the place of 1/5.

7.11 Independence of a pair of random variables

We again fix our probability space.

Definition 7.11.1 (Independence of a pair of random variables)**.** Let X, Y be random variables. We say that X and Y are **independent** if

$$(\forall u, v \in \mathbb{R})(\Pr(X = u \text{ and } Y = v) = \Pr(X = u)\cdot\Pr(Y = v)\,. \tag{7.38}$$

Exercise 7.11.2. If Y is almost constant (see Def. 7.9.1) then X and Y are independent.

Exercise* 7.11.3 (Independent implies uncorrelated)**.** If X, Y are independent then they are uncorrelated, i. e.,

$$\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)\,. \tag{7.39}$$

The next exercise asserts that the converse is false.

Exercise 7.11.4. Construct a probability space and two random variables that are uncorrelated but not independent. — Make sure you give a complete definition of your probability space: state the sample space and the probability distribution. Define your random variables by their table of values. Minimize the size of your sample space.

7.12 Independence of random variables

Definition 7.12.1. $X_1, \ldots, X_k$ are **independent** if

$$(\forall u_1, \ldots, u_k \in \mathbb{R}) \left(\Pr\left(X_1 = u_1, \ldots, X_k = u_k\right) = \prod_{i=1}^{k} \Pr(X_i = u_i \right)\,. \tag{7.40}$$

Exercise 7.12.2. Prove that the events $A_1, \ldots, A_k$ are independent if and only if their indicator variables are independent. – This is less obvious than it seems.

Exercise 7.12.3. Prove that the random variables $X_1, \ldots, X_k$ are independent if and only if for all choices of the numbers $u_1, \ldots, u_k$, the k events $X_1 = u_1, \ldots, X_k = u_k$ are independent. Show that this is also equivalent to the independence of all k-tuples of events of the form $X_1 < u_1, \ldots, X_k < u_k$.

Exercise 7.12.4. Prove: if $X_1, \ldots, X_k$ are independent then $f_1(X_1), \ldots, f_k(X_k)$ are also independent, where the f_i are arbitrary functions. For example, X_1^2, e^{X_2}, and $\cos(X_3)$ are independent.

Exercise 7.12.5. Prove: if X, Y, Z are independent random variables then $f(X, Y)$ and Z are also independent, where f is an arbitrary function. (For instance, $X + Y$ and Z, or XY and Z are independent.) Generalize this statement to several variables, grouped into blocks, and a function applied to each block.

Exercise 7.12.6. Let $X_1, \ldots, X_m$ be non-almost-constant random variables (see Def. 7.9.1) over a sample space of size n. Suppose the X_i are 4-wise independent (every four of them are independent). Prove: $n \geq \binom{m}{2}$. *Hint.* Prove that the $\binom{m}{2}$ random variables $X_i X_j$ $(1 \leq i < j \leq m)$ are linearly independent over $\mathbb{R}$ (as members of the space of functions $\Omega \to \mathbb{R}$). To prove linear independence, first prove that w.l.o.g. we may assume $(\forall i)(E(X_i) = 0)$; then use the "inner product" argument, using the function $E(ZY)$ in the role of an "inner product" of the random variables Z and Y.

Theorem 7.12.7 (Multiplicativity of the expected value)**.** *If $X_1, \ldots, X_m$ are independent, then*

$$\mathrm{E}(\prod_{i=1}^{m} X_i) = \prod_{i=1}^{m} \mathrm{E}(X_i). \tag{7.41}$$

Exercise 7.12.8. Prove this result for indicator variables.

Exercise 7.12.9. Prove: if X, Y are independent, then one can write X as a sum $X = c_1X_1 + \ldots + c_kX_k$ and Y as $Y = d_1Y_1 + \ldots + d_\ell Y_\ell$ where the X_i and Y_j are indicator variables and for every i, j, the variables X_i and Y_j are independent.

Exercise 7.12.10. Combine the two preceding exercises to a proof of the Theorem for $m = 2$ variables.

Exercise 7.12.11. Deduce the general case from the preceding exercise by induction on m, using Exercise 7.12.5.

This sequence completes the proof of Theorem 7.12.7. □

While this result required the full force of independence of our random variables, recall that the additivity of the variance only required pairwise independence. In fact even less, pairwise uncorrelatedness, suffices (Cor. 7.10.18).

Corollary 7.12.12. *Let $X_1, \ldots, X_n$ be random variables with the same standard deviation σ. Let us consider their average, $Y := (1/n)\sum_{i=1}^{n} X_i$. If the X_i are pairwise independent then $\sigma(Y) = \sigma/\sqrt{n}$.* □

Corollary 7.12.13 (Weak law of large numbers)**.** *Let $X_1, X_2, \ldots$ be an infinite sequence of pairwise independent random variables each with expected value m and standard deviation σ. Let $Y_n = (1/n)\sum_{i=1}^{n} X_i$. Then for any $\delta > 0$,*

$$\lim_{n\to\infty} \Pr(|Y_n - m| > \delta) = 0. \tag{7.42}$$

Proof: Use Chebyshev's inequality and the preceding corollary. We obtain that the probability in question is $\leq \sigma^2/(\delta n) \to 0$ (as $n \to \infty$). □

Remark 7.12.14. Strictly speaking, we bent our rules here. An infinite sequence of non-almost-constant, pairwise independent variables requires an infinite sample space. What we actually proved, then, is the following. Let us fix the values m and $\sigma \geq 0$. Assume that we are given an infinite sequence of finite probability spaces, and over the n^{th} space, we are given n independent random variables $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$. Let $Y_n = (1/n)\sum_{i=1}^{n} X_{n,i}$. Then for any $\delta > 0$, the limit relation (7.42) holds.

Exercise 7.12.15. You and the bank play the following game. You flip n coins; if X of them come up "Heads," you receive 2^X dollars.

1. You have to buy a ticket to play this game. What is the fair price of the ticket? *Hint:* it is the expected amount you will receive.

2. Prove: the probability that you break even (receive at least your ticket's worth) is exponentially small. *Hint:* At least how many "heads" do you need for you to break even?

3. Calculate the standard deviation of the variable 2^X. Your answer should be a simple formula. Evaluate it asymptotically; obtain an even simpler formula.

4. State what the "weak law of large numbers" would say for the variable 2^X. *Hint.* This law talks about the probability that 2^X is not within $(1\pm\epsilon)$-times its expectation.) Prove that the Law does NOT hold for this variable.

7.13 Strong concentration inequalities: the Bernstein–Hoeffding (Chernoff) bounds

Although the bound in the proof of the Weak Law of Large Numbers tends to zero, it does so rather slowly. If our variables are fully independent and bounded, much stronger estimates can be obtained by a method of **Sergey Bernstein** (1924, 1927, 1937) called the *moment generator function* method. Results derived by this method are often referred to as "Chernoff bounds," based on a 1952 paper by *Herman Chernoff* that rediscovered Bernstein's method (and did not derive those consequences attributed tom him). Another paper frequently referenced in this context is a 1963 paper by *Wassily Hoeffding* that is aware of Bernstein's work, and uses the moment generator method to derive several of those consequences often referred to as "Chernoff bounds." These bounds tend to be slightly stronger than the bounds found by Bernstein. We shall refer to them as the Bernstein–Hoeffding bounds.

The Bernstein–Hoeffding bounds go to zero exponentially fast as a function of n, and this is what most combinatorial applications require.

For example, let us consider a sequence of n independent coin flips; let X denote the number of heads in this sequence. Then $\mathrm{E}(X) = n/2$ and $\mathrm{Var}(X) = n/4$ (by the additivity of the variance). Therefore Chebyshev's inequality tells us that

$$\Pr(|X - n/2| \geq r\sqrt{n}) < \frac{1}{4r^2}. \tag{7.43}$$

Below we shall prove the much stronger inequality

$$\Pr(|X - n/2| \geq r\sqrt{n}) < 2\mathrm{e}^{-2r^2}. \tag{7.44}$$

under the same conditions.

The following corollary illustrates the power of inequality (7.44).

Corollary 7.13.1. *For any $\varepsilon > 0$, almost all graphs have no vertices of degree $< (1-\varepsilon)n/2$ or $> (1+\varepsilon)n/2$ where n is the number of vertices.*

Proof of the Corollary. Let $V = \{1, \ldots, n\}$ be the vertex set of our random graph. Let δ_i denote the degree of vertex i; so δ_i is the number of heads in a sequence of $(n-1)$ independent coin flips. Therefore, by inequality (7.44), we have that

$$\Pr(|\delta_i - (n-1)/2| \geq r\sqrt{n-1}) < 2\mathrm{e}^{-2r^2}. \tag{7.45}$$

Let us now set $r = \varepsilon\sqrt{n-1}$. Then we obtain

$$\Pr(|\delta_i - (n-1)/2| \geq \varepsilon(n-1)) < 2\mathrm{e}^{-2\varepsilon^2(n-1)}. \tag{7.46}$$

Therefore the probability that there exists an i such that $|\delta_i - (n-1)/2| \geq \varepsilon(n-1)$ is less than n times the right-hand side, i. e., less than $2n\mathrm{e}^{-2\varepsilon^2(n-1)}$. This quantity approaches zero at an exponential rate as $n \to \infty$.

The slight change in the statement (having changed n to $n-1$) can be compensated for by slightly reducing ε. □

Note that the same procedure using inequality (7.43) will fail. Indeed, setting $r = \varepsilon\sqrt{n-1}$ in inequality (7.43), the right-hand side will be $1/(4\varepsilon^2(n-1))$, and if we multiply this quantity by n, the result will be greater than 1 (if $\varepsilon < 1/2$), a meaningless upper bound for a probability.

Now we turn to the proof of inequality (7.44). Our discussion is based on the Appendix to the wonderful monograph

Noga Alon, Joel H. Spencer: "The Probabilistic Method."

It will be convenient to state the main result in terms of random variables with zero expected value.

Theorem 7.13.2 (Bernstein–Hoeffding bound for coin flips)**.** *Let X_i be independent random variables satisfying $\Pr(X_i = 1) = \Pr(X_i = -1) = 1/2$. Let $Y = \sum_{i=1}^n X_i$. Then for any $a > 0$,*

$$\Pr(Y \geq a) < \mathrm{e}^{-a^2/2n} \tag{7.47}$$

and

$$\Pr(|Y| \geq a) < 2\mathrm{e}^{-a^2/2n}. \tag{7.48}$$

Exercise 7.13.3. Deduce inequality (7.44) from this theorem.

Hint. Represent X as $\sum_{i=1}^n \vartheta_i$ where ϑ_i is the indicator variable of the i-th coin flip. Set $X_i = 2\vartheta_i - 1$ and $Y = \sum_{i=1}^n X_i$. Note that $X - n/2 = Y/2$. Apply Theorem 7.13.2 to the X_i and translate the result back to X.

Exercise 7.13.4. Prove that the following is true for almost all graphs $\mathcal{G}_n$ on n vertices: the degree of every vertex is within the interval $[0.49n, 0.51n]$. In answering this question, be sure to clearly state the meaning of each variable occurring in your formulas. Also pay close attention to the logical connectives ("and," "if-then," and quantifiers).

We now define the central concept of Bernstein's method.

Definition 7.13.5. Let Y be a random variable. The function $m_Y(t) = \mathrm{E}(\mathrm{e}^{tY})$ is called the **moment generator function** of Y.

Now we turn to the proof of Theorem 7.13.2.

Let t be a positive real number. A specific value will be assigned to t later. Let us consider the random variables $Z_i := \exp(tX_i)$. (Notation: $\exp(x) = \mathrm{e}^x$.) The Z_i are again independent (for any fixed t) by Exercise 7.12.4. Therefore we can apply the multiplicativity of the expected value to them to calculate the moment generator function of Y:

$$\mathrm{E}(\mathrm{e}^{tY}) = \mathrm{E}(\exp(\sum_{i=1}^{n} tX_i)) = \mathrm{E}(\prod_{i=1}^{n} Z_i) = \prod_{i=1}^{n} \mathrm{E}(Z_i) = \prod_{i=1}^{n} \mathrm{E}(\exp(tX_i)). \tag{7.49}$$

Applying Markov's inequality to the variable e^{tY}, we conclude that

$$\Pr(Y \geq a) = \Pr(\mathrm{e}^{tY} \geq \mathrm{e}^{ta}) \leq \prod_{i=1}^{n} \mathrm{E}(\exp(tX_i))\mathrm{e}^{-ta}. \tag{7.50}$$

Recall the definition of the hyperbolic cosine function, $\cosh(x) = (\mathrm{e}^x + \mathrm{e}^{-x})/2$. Observe that

$$\mathrm{E}(\exp(tX_i)) = \cosh(t). \tag{7.51}$$

Therefore the preceding inequality implies that

$$\Pr(Y \geq a) < \frac{\cosh(t)^n}{\mathrm{e}^{ta}}. \tag{7.52}$$

This is true for every $t > 0$. All we need to do is choose t appropriately to obtain the strongest possible result. To this end we need the following simple observation.

Lemma 7.13.6. *For all real numbers x,*

$$\cosh(x) \leq \mathrm{e}^{x^2/2}.$$

Proof: Compare the Maclaurin series of the two sides. On the left-hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \dots \tag{7.53}$$

On the right-hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} = 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \dots . \tag{7.54}$$

Comparing the denominators in the corresponding terms, we see that $(2k)! \geq 2^k k!$ for all $k \geq 0$ (why?), so the right-hand side dominates the left-hand side term by term. □

Consequently, from inequality (7.52) we infer that

$$\Pr(Y \geq a) < \exp(t^2 n/2 - ta). \tag{7.55}$$

The expression $t^2 n/2 - ta$ is minimized when $t = a/n$; setting $t := a/n$ we conclude that $\Pr(Y \geq a) < \exp(-a^2/2n)$, as required.

Replacing each X_i by $-X_i$ we obtain the inequality $\Pr(Y \leq -a) < \exp(-a^2/2n)$; adding this to the preceding inequality we obtain $\Pr(|Y| \geq a) < 2\exp(-a^2/2n)$. □

We note that this technique works under much more general circumstances. We state a useful and rather general case, noting that even this result does not exploit the full power of the method.

Theorem 7.13.7 (Bernstein–Hoeffding bound for bounded variables)**.** *Let X_i be independent random variables satisfying $|X_i| \leq 1$ and $\mathrm{E}(X_i) = 0$. Let $Y = \sum_{i=1}^{n} X_i$. Then for any $a > 0$,*

$$\Pr(Y \geq a) < \mathrm{e}^{-a^2/2n} \tag{7.56}$$

and

$$\Pr(|Y| \geq a) < 2\mathrm{e}^{-a^2/2n}. \tag{7.57}$$

Proof: Fix a value $t > 0$. Let

$$h_t(x) = \cosh(t) + x \cdot \sinh(t). \tag{7.58}$$

(Recall that $\sinh(t) = (\mathrm{e}^t - \mathrm{e}^{-t})/2$ is the hyperbolic sine function.) This is a linear function of x. Observe that $h_t(x) \geq \mathrm{e}^{tx}$ for all x in the interval $-1 \leq x \leq 1$. (The graph of $h_t(x)$ over the interval $[-1, 1]$ is the segment connecting the corresponding two points of the graph of the function e^{tx}, and e^{tx} is a convex function.)

Moreover, because of the linearity of the $h_t(x)$ function, we have $\mathrm{E}(h_t(X_i)) = h_t(\mathrm{E}(X_i)) = h_t(0) = \cosh(t)$. Therefore

$$\mathrm{E}(\mathrm{e}^{tX_i}) \leq \mathrm{E}(h_t(X_i)) = \cosh(t). \tag{7.59}$$

From here on the proof is identical with the proof of Theorem 7.13.2. As before, we set $t = a/n$. □

Exercise 7.13.8. Prove: for almost all graphs $G = (V, E)$ with n vertices,

$$(\forall x \in V)(0.49n < \deg(x) < 0.51n)\,. \tag{7.60}$$

In other words, if p_n denotes the probability of the event described in Eq. (7.60) then $\lim_{n\to\infty} p_n = 1$.

Explain, why this result does not follow from Chebyshev's inequality.

Exercise 7.13.9. A vertex z is a *common neighbor* of vertices x and y in a graph G if both x and y are adjacent to z in G. Let $N(x, y)$ denote the number of common neighbors of x and y. Prove that the following statement is true for *almost all* graphs $G = (V, E)$ with n vertices:

$$(\forall x \neq y \in V)(0.24n < N(x, y) < 0.26n)\,. \tag{7.61}$$

Exercise 7.13.10. Determine (a) the expected number and (b) the variance of the number of edges of