Discrete Mathematics
Lecture Notes
Incomplete Preliminary Version

Instructor: László Babai

Last revision: June 22, 2003
Last update: January 5, 2023

Contents

List of Figures

Chapter 1

Logic

1.1 Quantifier notation

Quantifier notation: $\forall$ - "universal quantifier," $\exists$ - "existential quatifier."

$(\forall x)$ is read as "for all x"

$(\exists x)$ is read as "there exists x **such that**"

$(\forall x, \text{statement}(x))$ is read as "for all x such that statement(x) holds,. . . "

Example. $(\forall x \neq 0)(\exists y)(xy = 1)$ says that every x other than zero has a multiplicative inverse. The validity of this statement depends on the universe over which the variables range. The statement holds (is true) over $\mathbb{R}$ (real numbers) and $\mathbb{Q}$ (rational numbers) but does not hold over $\mathbb{Z}$ (integers) or $\mathbb{N}$ (nonnegative integers). It holds over $\mathbb{Z}_m$ (the set of residue classes modulo m) if m is prime but not if m is composite. (Why?)

1.2 Problems

Several of the problems below will refer to the *divisibility* relation between integers.

Definition 1.2.1. Let a, b be integers. We say that $a \mid b$ ("a divides b") if $(\exists x)(ax = b)$. (The universe of the quantifiers is $\mathbb{Z}$, the set of integers (positive, negative, zero).)

From this definition we see that $7 \mid 21$ (because $x = 3$ satisfies $7x = 21$); $5 \mid -5$ (because $x = -1$ satisfies $5x = -5$); $0 \mid 0$ (because $x = 17$ (or any other x) satisfies $0x = 0$).

Does our conclusion $0 \mid 0$ violate the prohibition agains division by zero? By no means; division by zero continues to be a no-no. But read the definition of divisibility: it involves *multiplication*, not division. Nothing can stop us from *multiplying* a number by zero.

Remark. Most (but not all) Discrete Mathematics texts deliberately misstate the definition of divisibility to exclude $0 \mid 0$ from the definition. This abomination stems from many textbook authors' contempt for their readers' intelligence; the result is a multitude of unnecessary case distinctions, destroying a fundamental element of mathematical aesthetics. (To these authors, for instance, $x \mid x$ does not hold for all x; there is an exception: $x = 0$. And then, to them, $x-y$ does not always divide x^2-y^2; to them, the cases when $x = y$ are exceptions.) We do not follow this deplorable textbook trend; to us (as well as tpo any mathematician), $(\forall x)(x \mid x)$ and $(\forall x)(\forall y)(x - y \mid x^2 - y^2)$.

Exercise 1.2.2. Restate the following statements in plain English and prove them. The universe is $\mathbb{Z}$.

(a) $(\forall x)(x \mid x)$. In particular, $0 \mid 0$.

(b) $(\forall x)(\forall y)(x - y \mid x^2 - y^2)$.

(c) $(\forall x)(1 \mid x)$.

(d) $(\forall x)(x \mid 0)$.

(e) $(\forall x)(\text{ if } (\forall y)(x \mid y) \text{ then } x = \pm 1)$.

(f) $(\forall x)(\text{ if } (\forall y)(y \mid x) \text{ then } x = 0)$.

Definition 1.2.3. (Congruence) Let a, b, m be integers. We say that $a \equiv b \pmod{m}$ ("a is congruent to b modulo m") if $m \mid a - b$.

Examples: $11 \equiv -10 \pmod{-7}$ because $-7 \mid 11 - (-10) = 21$. Two integers are congruent modulo 2 exactly if they have the same parity (both are even or both are odd).

Exercise 1.2.4. Prove the following statements. The universe is $\mathbb{Z}$.

(a) $(\forall x)((\forall y)(\forall z)(y \equiv z \pmod{x}) \iff x = \pm 1)$.

(b) $(\forall x)(\forall y)(x \equiv y \pmod{0} \quad \iff \quad x = y)$.

(c) $(\forall x \neq \pm 1)(\forall y)(\exists z)(y \not\equiv z \pmod{x})$.

Exercise 1.2.5. Decide whether each of the following statements is true or false. *State* and *prove* your answers. In these statements, the universe for the variables x, y, k is $\mathbb{Z}$, the set of *integers*. Warning: in interpreting the formulas, *the order of the quantifiers matters!* $(\forall x)(\forall y)(P(x, y))$ is the same as $(\forall y)(\forall x)(P(x, y))$; $(\exists x)(\exists y)(P(x, y))$ is the same as $(\exists y)(\exists x)(P(x, y))$; but $(\forall x)(\exists y)(P(x, y))$ is NOT the same as $(\exists y)(\forall x)(P(x, y))$!

(a) $(\forall x)(\forall y)(x + y \mid x^2 - y^2)$.

(b) $(\forall x)(\forall y)(x + y \mid x^2 + y^2)$.

(c) $(\exists x)(\forall y)(x + y \mid x^2 + y^2)$.

(d) $(\forall x)(\exists y)(x^2 + y^2 \equiv 1 \pmod{x + y})$.

(e) $(\forall x)(\forall y)(\forall k)$ (if $k \geq 1$ then $x^k \equiv y^k \pmod{x - y}$).

(f) $(\forall x)(\exists y)(x \neq y$ and $x \mid y$ and $x \equiv y \pmod 7)$.

(g) $(\exists y)(\forall x)(x \neq y$ and $x \mid y$ and $x \equiv y \pmod 7)$.

(h) $(\forall x)(\forall y)($ if $x \mid y$ and $x \neq y$ then $x < y)$.

Exercise 1.2.6. True or false (prove your answer):

$$(\forall x)(\exists y)(\forall z)((x - 5y)z \not\equiv 1 \pmod{17}).$$

(The universe of the variables is the set of integers.)

Negation of quantified formulas. If A is a statement then $\neg A$ denotes its negation; so $\neg A$ is true if and only if A is false. $\Leftrightarrow$ denotes logical equivalence ("if and only if").

Exercise 1.2.7. Let $P(x)$ be a statement in variable x.

(a) Prove: $\neg(\forall x)(P(x)) \Leftrightarrow (\exists x)(\neg P(x))$.

(b) Prove: $\neg(\exists x)(P(x)) \Leftrightarrow (\forall x)(\neg P(x))$.

(c) Let $Q(x, y)$ be a statement in two variables. Prove: $\neg(\forall x)(\exists y)(Q(x, y)) \Leftrightarrow (\exists x)(\forall y)(\neg Q(x, y))$.

Exercise 1.2.8. Let $P(x, y)$ be a statement about the variables x and y. Consider the following two statements: $A := (\forall x)(\exists y)(P(x, y))$ and $B := (\exists y)(\forall x)(P(x, y))$. The universe is the set of integers.

(a) Prove: $(\forall P)(B \Rightarrow A)$ ("B always implies A," i. e., for all P, if B is true then A is true).

(b) Prove: $\neg(\forall P)(A \Rightarrow B)$ (i. e., A does not necessarily imply B). In other words, $(\exists P)(A \not\Rightarrow B)$. To prove this, you need to construct a counterexample, i. e., a statement $P(x, y)$ such that the corresponding statement A is true but B is false. Make $P(x, y)$ as simple as possible. *Hint.* Three symbols suffice. These include x and y.

Quantifier alternation and games.

Exercise 1.2.9. Digest and generalize the following. Consider a chess-puzzle which says "white moves and wins in 2 moves." Let $W(x)$ denote the statement that the move x is available to White; and $B(x, y)$ that the move y is available to Black after White's move x; and $W(x, y, z)$ the statement that move z is avaliable to White after White moved x and Black moved y. Let $C(x, y, z)$ denote the statement that after moves x, y, z, Black is checkmated. Now the puzzle's claim can be formalized in the following quantified formula:

$$(\exists x, W(x))(\forall y, B(x, y))(\exists z, W(x, y, z))(C(x, y, z)).$$

Chapter 2

Asymptotic Notation

Last update: January 5, 2023

2.1 Limit of sequence

Notation: $\exp(x) = \mathrm{e}^x$.

In combinatorial contexts, the symbol $[n]$ will be used to denote $\{1, 2, \dots, n\}$.

Definition 2.1.1 (finite limit of a sequence)**.** Let $\{a_n\}$ be a sequence of real or complex numbers. We write $\lim_{n\to\infty} a_n = c$ (or simply $a_n \to c$) if

$$(\forall \epsilon > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(|a_n - c| \leq \epsilon).$$

We say that a sequence *converges* if it has a finite limit.

Definition 2.1.2 (infinite limit of a sequence)**.** Let a_n be a sequence of real or complex numbers. We write $\lim_{n\to\infty} a_n = \infty$ (or simply $a_n \to \infty$) if

$$(\forall L)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(a_n \geq L).$$

Exercise 2.1.3. What is $\lim_{n\to\infty}(1 + x/n)^n$?

Exercise 2.1.4. (a) Consider the sequence $\{a_n\}$ defined by the recurrence $a_{n+1} = \sqrt{2}^{a_n}$ with the initial condition $a_0 = 1$. Prove that $\lim_{n\to\infty} a_n$ exists; find the limit.

(b) Prove that the previous statement becomes false if we replace $\sqrt{2}$ by 1.5. What is the largest number (in place of $\sqrt{2}$) for which the sequence converges?

2.2 Asymptotic Equality and Inequality

THIS CHAPER HAS BEEN SUPESRSEDED. SEE SEPARATE SET OF NOTES UNDER THE SAME TITLE

Often, we are interested in comparing the rate of growth of two functions, as inputs increase in length. Asymptotic equality is one formalization of the idea of two functions having the "same rate of growth."

Definition 2.2.1. We say a_n is *asymptotically equal* to b_n (denoted $a_n \sim b_n$) if $\lim_{n\to\infty} a_n/b_n = 1$. For the purposes of this definition, we set $0/0 = 1$.

Observation. If $c \neq 0$ is a constant then the statement $a_n \sim c$ (where c means the sequence $c, c, \dots$) is equivalent to $a_n \to c$ (where c means the number c).

Last update: January 5, 2023

Exercise 2.2.2. Prove: $a_n \sim 0$ if and only if $(\exists n_0)(\forall n \geq n_0)(a_n = 0)$, i. e., $a_n = 0$ for all sufficiently large n.

Exercise 2.2.3. Let $\mathcal{S}$ denote the set of sequences of real or complex numbers. Prove that $\sim$ is an *equivalence relation* on $\mathcal{S}$, i. e., the relation "$\sim$" is

(a) *reflexive:* $a_n \sim a_n$;

(b) *symmetric:* if $a_n \sim b_n$ then $b_n \sim a_n$; and

(c) *transitive:* if $a_n \sim b_n$ and $b_n \sim c_n$ then $a_n \sim c_n$.

Exercise 2.2.4. Prove: if $a_n \sim b_n$ and $c_n \sim d_n$ then $a_n c_n \sim b_n d_n$. If, moreover, $c_n d_n \neq 0$ for all sufficiently large n then $a_n/c_n \sim b_n/d_n$. (Note that a finite number of undefined terms do not invalidate a limit relation.)

Exercise 2.2.5. Consider the following statement.

$$\text{If } a_n \sim b_n \text{ and } c_n \sim d_n \text{ then } a_n + c_n \sim b_n + d_n. \tag{2.1}$$

1. Prove that (2.1) is false.

2. Prove: if $a_n c_n > 0$ then (2.1) is true. *Hint.* Prove: if $a, b, c, d > 0$ and $a/b < c/d$ then $a/b < (a+c)/(b+d) < c/d$.

Exercise 2.2.6.

1. If $f(x)$ and $g(x)$ are polynomials with respective leading terms ax^n and bx^m then $f(n)/g(n) \sim (a/b)x^{n-m}$.

2. $\sin(1/n) \sim \ln(1 + 1/n) \sim 1/n$.

3. $\sqrt{n^2+1} - n \sim 1/2n$.

4. If f is a function, differentiable at zero, $f(0) = 0$, and $f'(0) \neq 0$, then $f(1/n) \sim f'(0)/n$. See that items 2 and 3 in this exercise follow from this.

Exercise 2.2.7. Find two sequences of positive real numbers, $\{a_n\}$ and $\{b_n\}$, such that $a_n \sim b_n$ but $a_n^n \not\sim b_n^n$.

Next we state some of the most important asymptotic formulas in mathematics.

Theorem 2.2.8 (Stirling's Formula)**.**

$$n! \sim \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}.$$

Exercise 2.2.9. Prove: $\displaystyle\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}$.

Exercise 2.2.10. Give a very simple proof, without using Stirling's formula, that $\ln(n!) \sim n \ln n$.

Theorem 2.2.11 (The Prime Number Theorem)**.** *Let $\pi(x)$ be the number of primes less than or equal to x.*

$$\pi(x) \sim \frac{x}{\ln x},$$

where ln *denotes the natural logarithm function.*

Exercise 2.2.12. Let p_n be the n-th prime number. Prove, using the Prime Number Theorem, that $p_n \sim n \ln n$.

Exercise 2.2.13. *Feasibility of generating random prime numbers.* Estimate, how many random ≤ 100-digit integers should we expect to pick before we encounter a prime number? (We generate our numbers by choosing the 100 digits independently at random (initial zeros are permitted), so each of the 10^{100} numbers has the same probability to be chosen.) Interpret this question as asking the reciprocal of the probability that a randomly chosen integer is prime.

Definition 2.2.14. A *partition* of a positive integer n is a representation of n as a sum of positive integers: $n = x_1 + \cdots + x_k$ where $x_1 \le \cdots \le x_k$. Let $p(n)$ denote the number of partitions of n.

Examples: $p(1) = 1$, $p(2) = 2$, $p(3) = 3$, $p(4) = 5$. The 5 representations of 4 are $4 = 4$; $4 = 1 + 3$; $4 = 2 + 2$; $4 = 1 + 1 + 2$; $4 = 1 + 1 + 1 + 1$. One of the most amazing asymptotic formulas in discrete mathematics gives the growth of $p(n)$.

Theorem 2.2.15 (Hardy-Ramanujan Formula)**.**

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\frac{2\pi}{\sqrt{6}}\sqrt{n}\right). \tag{2.2}$$

Definition 2.2.16. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. We say that a_n is *greater than or asymptotically equal to* b_n, denoted as $a_n \gtrsim b_n$ if $a_n \sim \max\{a_n, b_n\}$.

Exercise 2.2.17. Prove: $a_n \gtrsim b_n$ if and only if $b_n \sim \min\{a_n, b_n\}$.

Exercise 2.2.18. Prove: if $a_n \sim b_n$ then $a_n \gtrsim b_n$.

Exercise 2.2.19. Prove: if $a_n \gtrsim b_n$ and $b_n \gtrsim a_n$ then $a_n \sim b_n$.

Exercise 2.2.20. Prove: if $a_n \gtrsim b_n$ and $b_n \gtrsim c_n$ then $a_n \gtrsim c_n$.

Exercise 2.2.21. Conclude from the preceding exercises that the "$\gtrsim$" relation is a partial order on the set of asymptotic equivalence classes of sequences of real numbers.

Exercise 2.2.22. Prove: $a_n \gtrsim 0$ if and only if $(\exists n_0)(\forall n \ge n_0)(a_n \ge 0)$, i. e., $a_n \ge 0$ for all sufficiently large n.

Exercise 2.2.23. Prove: if $a_n \gtrsim b_n \geq 0$ and $c_n \gtrsim d_n \geq 0$ then $a_n + c_n \gtrsim b_n + d_n$.

Exercise 2.2.24. (a) Let $a_n, b_n \geq 0$. Prove that $a_n \gtrsim b_n$ if and only if $(\forall \epsilon > 0)(\exists n_0)(\forall n > n_0)(a_n \geq b_n(1 - \epsilon))$.

(b) Show that the same formula does not define the relation "$a_n \gtrsim b_n$" if we omit the condition $a_n, b_n \geq 0$.

Exercise 2.2.25. Assume $b_n \to \infty$ and $a_n \geq b_n^2 \ln b_n$. Prove: $b_n \lesssim c\sqrt{a_n / \ln a_n}$, where c is a constant. Determine the smallest value of c for which this statement follows from the assumptions.

2.3 Little-oh and little-omega notation

Definition 2.3.1. We say that $a_n = o(b_n)$ ("a_n is little oh of b_n") if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Observation. So $a_n = o(1)$ means $\lim_{n\to\infty} a_n = 0$.

Exercise 2.3.2. Show: if $a_n = o(c_n)$ and $b_n = o(c_n)$ then $a_n \pm b_n = o(c_n)$.

Exercise 2.3.3. Consider the following statement:

$$\text{If } a_n = o(b_n) \text{ and } c_n = o(d_n) \text{ then } a_n + c_n = o(b_n + d_n). \tag{2.3}$$

1. Show that statement (2.3) is false.

2. Prove that statement (2.3) becomes true if we assume $b_n, d_n > 0$.

Exercise 2.3.4. Show that $a_n \sim b_n \iff a_n = b_n(1 + o(1))$.

Exercise 2.3.5. Use the preceding exercise to give a second proof of (2.1) when $a_n, b_n, c_n, d_n > 0$.

Exercise 2.3.6. Construct sequences $a_n, b_n > 1$ such that $a_n = o(b_n)$ and $\ln a_n \sim \ln b_n$.

Exercise 2.3.7. Let $a_n, b_n > 1$. (a) Prove that the relation $a_n = o(b_n)$ does NOT follow from the relation $\ln a_n = o(\ln b_n)$. (b) If we additionally assume that $b_n \to \infty$ then $a_n = o(b_n)$ DOES follow from $\ln a_n = o(\ln b_n)$.

Definition 2.3.8. We say that $a_n = \omega(b_n)$ ("a_n is little omega of b_n") if $b_n = o(a_n)$.

2.4 Big-Oh, Omega, Theta notation (O, Ω, Θ)

Definition 2.4.1. We say that

1. $a_n = O(b_n)$ (a_n is "big oh" of b_n) if $|a_n/b_n|$ is bounded (0/0 counts as "bounded"), i. e.,

$$(\exists C > 0, n_0 \in \mathbb{N})(\forall n > n_0)(|a_n| \leq C|b_n|).$$

2. $a_n = \Omega(b_n)$ if $b_n = O(a_n)$, i. e., if $|b_n/a_n|$ is bounded $(\exists c > 0, n_0 \in \mathbb{N})(\forall n > n_0)(|a_n| \geq c|b_n|)$

3. $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$, i. e.,

$$(\exists C, c > 0, n_0 \in \mathbb{N})(\forall n > n_0)(c|b_n| \leq |a_n| \leq C|b_n|).$$

Exercise 2.4.2. Suppose the finite or infinite limit $\lim_{n\to\infty} |a_n/b_n| = L$ exists. Then

(a) $a_n = o(b_n)$ if and only if $L = 0$; and

(b) $a_n = \Theta(b_n)$ if and only if $0 < L < \infty$.

(c) $a_n = \omega(b_n)$ if and only if $L = \infty$;

Exercise 2.4.3. Construct sequences $a_n, b_n > 0$ such that $a_n = \Theta(b_n)$ but the limit $\lim\limits_{n\to\infty} a_n/b_n$ does not exist.

Exercise 2.4.4. Let $a_n, b_n > 0$. Show: $a_n = \Theta(b_n) \iff \ln a_n = \ln b_n + O(1)$.

Exercise 2.4.5. Show: if $a_n = O(c_n)$ and $b_n = O(c_n)$ then $a_n + b_n = O(c_n)$.

Exercise 2.4.6. Consider the statement "if $a_n = \Omega(c_n)$ and $b_n = \Omega(c_n)$ then $a_n + b_n = \Omega(c_n)$. (a) Show that this statement is false. (b) Show that if we additionally assume $a_n b_n > 0$ then the statement becomes true.

Exercise 2.4.7. Let $a_n, b_n > 1$. Suppose $a_n = \Theta(b_n)$. Does it follow that $\ln a_n \sim \ln b_n$?

1. Show that even $\ln a_n = \Omega(\ln b_n)$ does not follow.

2. Show that if $a_n \to \infty$ then $\ln a_n \sim \ln b_n$ follows.

Exercise 2.4.8. Let $a_n, b_n > 1$. Suppose $a_n = \Omega(b_n)$. Does it follow that $\ln a_n \gtrsim \ln b_n$?

1. Show that even $\ln a_n = \Omega(\ln b_n)$ does not follow.

2. Show that if $a_n \to \infty$ then $\ln a_n \gtrsim \ln b_n$ follows.

Exercise 2.4.9. Let $a_n, b_n > 0$. Consider the relations

(A) $a_n = O(2^{b_n})$ and (B) $a_n = 2^{O(b_n)}$.

(a) Prove: the relation (B) does NOT follow from (A).

(b) Prove: if $a_n > 0.01$ and $b_n > 0.01$ then (B) DOES follow from (A).

Note. $a_n = 2^{O(b_n)}$ means that $a_n = 2^{c_n}$ where $c_n = O(b_n)$.

Exercise 2.4.10. Prove: if $a_n = \Omega(b_n)$ and $a_n = \Omega(c_n)$ then $a_n = \Omega(b_n + c_n)$.

Note. We say that the "statement A *implies* statement B" if B follows from A.

Exercise 2.4.11. (a) Prove that the relations $a_n = O(b_n)$ and $a_n = O(c_n)$ do NOT imply $a_n = O(b_n + c_n)$.

(b) Prove that if $a_n, b_n > 0$ then the relations $a_n = O(b_n)$ and $a_n = O(c_n)$ DO imply $a_n = O(b_n + c_n)$.

Exercise 2.4.12. Prove: $\sum_{i=1}^{n} 1/i = \ln n + O(1)$.

2.5 Prime Numbers

Exercise$^+$ 2.5.1. Let $P(x)$ denote the product of all prime numbers $\le x$. Consider the following statement: $\ln P(x) \sim x$. Prove that this statement is equivalent to the Prime Number Theorem.

Exercise$^+$ 2.5.2. Prove, without using the Prime Number Theorem, that

$$\ln P(x) = \Theta(x).$$

Hint. For the easy upper bound, observe that the binomial coefficient $\binom{2n}{n}$ is divisible by the integer $P(2n)/P(n)$. This observation yields $P(x) \le 4^x$. For the lower bound, prove that if a prime power p^t divides the binomial coefficient $\binom{n}{k}$ then $p^t \le n$. From this it follows that $\binom{2n}{n}$ divides the product $P(2n)P((2n)^{1/2})P((2n)^{1/3})P((2n)^{1/4})\dots$. Use the upper bound to estimate all but the first term in this product.

2.6 Partitions

Exercise 2.6.1. Let $p(n,k)$ denote the number of those partitions of n which have at most k terms. Let $q(n,k)$ denote the number of those partitions in which every term is $\le k$. Observe that $p(n,1) = q(n,1) = 1$ and $p(n,n) = q(n,n) = p(n)$. (Do!) Let $\widetilde{p}(n) = \sum_{i=0}^{n} p(i)$ and let $\widetilde{p}(n,k) = \sum_{i=0}^{n} p(i,k)$.

1. Prove: $p(n,k) = q(n,k)$.

2. Compute $p(n,2)$. Give a very simple formula.

3. Compute $p(n,3)$. Give a simple formula.

4. Prove: $\widetilde{p}(n) \le \widetilde{p}(n,k)^2$, where $k = \lfloor\sqrt{n}\rfloor$. *Hint.* Use part 1 of this exercise.

Exercise 2.6.2. Using the notation proved in Exercise 2.6.1, prove the following.

(a) $\widetilde{p}(n,k) < \binom{n+k}{k}$

(b) $\log p(n) = O(\sqrt{n}\log n)$. *Hint.* Use (a) and part 4 of Exercise 2.6.1.

Exercise$^+$ 2.6.3. Prove, without using the Hardy–Ramanujan formula, that

$$\ln p(n) = \Theta(\sqrt{n}).$$

Hint. $\ln p(n) = \Omega(\sqrt{n})$ is easy (2 lines). The upper bound is harder. Use the preceding exercise, especially item 4. When estimating $p(n,\sqrt{n})$, split the terms of your partition into sets $\{x_i \le \sqrt{n}\}$, $\{\sqrt{n} < x_i \le 2\sqrt{n}\}$, $\{2\sqrt{n} < x_i \le 4\sqrt{n}\}$, $\{4\sqrt{n} < x_i \le 8\sqrt{n}\}$, etc.

Exercise$^+$ 2.6.4. Let $p'(n)$ denote the number of partitions of n such that all terms are primes or 1. Example: $16 = 1+1+1+3+3+7$. Prove:

$$\ln p'(n) = \Theta\left(\sqrt{\frac{n}{\ln n}}\right).$$

Exercise 2.6.5. Let $r(n)$ denote the number of different integers of the form $\prod x_i!$ where $x_i \ge 1$ and $\sum x_i = n$. (The x_i are integers.) Prove:

$$p'(n) \le r(n) \le p(n).$$

OPEN QUESTIONS. Is $\log r(n) = \Theta(\sqrt{n})$? Or perhaps, $\log r(n) = \Theta(\sqrt{n/\log n})$? Or maybe $\log r(n)$ lies somewhere between these bounds?

2.7 Problems

Exercise 2.7.1.

1. (1 point) Describe in words what it means for a sequence a_n that $a_n = O(1)$ (big-Oh of 1).

2. (2 points) Suppose $a_n = O(1)$. Does it follow that the sequence a_n has a limit? (Prove your answer.)

3. (2 points) Suppose the sequence a_n has a finite limit. Does it follow that $a_n = O(1)$? Prove your answer.

Exercise 2.7.2. Let $a_n, b_n > 1$. True or false: if $a_n \sim b_n$ then $a_n^n = \Theta(b_n^n)$. Prove your answer.

Exercise 2.7.3. Prove: if $a_n, b_n, c_n, d_n > 0$ and $a_n = O(b_n)$ and $c_n = O(d_n)$ then $a_n + c_n = O(b_n + d_n)$. State the constant implicit in the conclusion as a function of the constants implicit in the conditions.

Exercise 2.7.4. Using the fact that $\ln x = o(x)$, prove that $(\ln y)^{100} = o(\sqrt{y})$. $(x, y \to \infty.)$ Do not use calculus.

Exercise 2.7.5. True or false (prove your answer):

$$2^{\binom{n}{2}} \sim 2^{n^2/2}.$$

Exercise 2.7.6. Construct two sequences, $\{a_n\}$ and $\{b_n\}$ such that $a_n > 1$, $b_n > 1$, $a_n \sim b_n$, and $a_n^n = o(b_n^n)$.

Exercise 2.7.7. Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers. Prove: if $a_n \to \infty$ and $a_n = \Theta(b_n)$ then $\ln(a_n) \sim \ln(b_n)$.

Exercise 2.7.8. Recall that a sequence $\{a_n\}$ is *polynomially bounded* if $(\exists C)(a_n = O(n^C))$. Decide whether or not each of the following sequences is polynomialy bounded. Prove your answers.

1. $n^3 \ln(n^2 + 5)$
2. $5^{\ln n}$
3. $\lfloor \ln n \rfloor !$

Exercise 2.7.9. Construct two sequences, $\{a_n\}$ and $\{b_n\}$ such that $a_n > 1$, $b_n > 1$, $a_n \sim b_n$, and $a_n^n = o(b_n^n)$.

Exercise 2.7.10. Let $f_n = (1 + 1/\sqrt{n})^n$ and $g_n = \mathrm{e}^{\sqrt{n}}$. Prove: $f_n = \Theta(g_n)$ but $f_n \not\sim g_n$. Show that in fact $\lim_{n\to\infty} f_n/g_n = 1/\sqrt{\mathrm{e}}$.

Exercise 2.7.11. Consider the statement

$\lim x^y = 1$ is "almost always true" as $x, y \to 0^+$.

Give a definition of "almost always" in this context, then prove the statement.

Exercise 2.7.12. Let $\{a_n\}$ be a sequence of positive integers, and assume $a_n \to \infty$. Let $b_n = \binom{a_n}{3}$. Prove that $a_n \sim c \cdot b_n^d$ for some constants c, d. Determine the values of c and d.

Chapter 3

Convex Functions and Jensen's Inequality

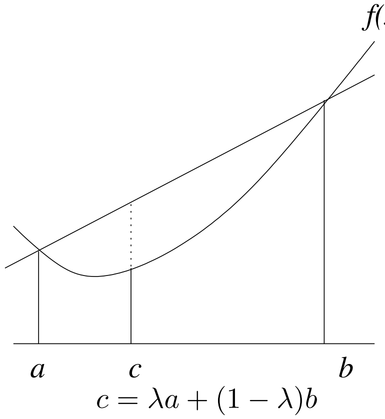


Figure 3.1: Definition of convexity

Definition 3.1.1. Let $f(x)$ be a real function defined over a finite or infinite interval. We say that $f(x)$ is a *convex* function if for all a, b in its domain and all real numbers λ in the interval $0 \leq \lambda \leq 1$, the inequality

$$f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda) f(b)$$

holds. The function $g(x)$ is *concave* if $-g(x)$ is convex. See Figure 3.

Exercise 3.1.2. Prove the following sufficient condition of convexity: If $f(x)$ is twice differentiable and its second derivative is always ≥ 0 then $f(x)$ is convex.

Exercise 3.1.3. Prove the following sufficient condition of convexity: If $f(x)$ is continuous and the inequality $f\left(\frac{a+b}{2}\right) \leq \frac{f(a)+f(b)}{2}$ holds for all a, b in its domain then $f(x)$ is convex.

Exercise 3.1.4. (a) The functions x^2, $\binom{x}{2}$, e^x are convex. (b) The functions $\sqrt{x}$, $\ln x$ are concave. (c) The function $\sin x$ is concave over the interval $[0, \pi]$ and convex over the interval $[\pi, 2\pi]$.

Exercise 3.1.5. (a) A continuous convex function is *unimodal:* it decreases to its minimum and then it increases. (b) If a continuous convex function is invertible then it is monotone (increasing or decreasing). (c) The inverse of a monotone increasing continuous convex function is concave. (d) The inverse of a monotone decreasing convex function is convex.

Theorem 3.1.6 (Jensen's Inequality). *If $f(x)$ is a convex function then for any choice of real numbers $x_1, \dots, x_k$ from the domain of f,*

$$f\left(\frac{\sum_{i=1}^k x_i}{k}\right) \leq \frac{\sum_{i=1}^k f(x_i)}{k}.$$

Exercise 3.1.7. Prove Jensen's Ineqality. *Hint.* Induction on k.

Exercise 3.1.8. Prove the inequality between the **arithmetic and quadratic means:** for all real $x_1, \dots, x_k$,

$$\frac{x_1 + \dots + x_k}{k} \leq \sqrt{\frac{x_1^2 + \dots + x_k^2}{k}}.$$

Hint 1. Use the convexity of $f(x) = x^2$ and Jensen's Inequality.
Hint 2. Give a 1-line proof using the Cauchy–Schwarz Inequality.
Hint 3. Give a simple direct proof (do not use either Jensen's Inequality or Cauchy–Schwarz).

Exercise 3.1.9. In the proof of the Kővári–Sós–Turán theorem (Exercise 6.1.41), we applied Jensen's Inequality to $f(x) = \binom{x}{2} = x(x-1)/2$. Modify the proof so that Jensen's Inequality is avoided and the inequality between the arithmetic and quadratic means is used instead.

Exercise 3.1.10. Prove the inequality between the **arithmetic and geometric means:** if $x_1, \dots, x_k > 0$ then

$$\frac{x_1 + \dots + x_k}{k} \geq (x_1 x_2 \dots x_k)^{1/k}.$$

Hint. Use the concavity of the natural logarithm function, ln.

Chapter 4

Basic Number Theory

4.1 Introductory Problems: g.c.d., congruences, multiplicative inverse, Chinese Remainder Theorem, Fermat's Little Theorem

Notation: Unless otherwise stated, all variables in this chapter are *integers.* For $n \geq 0$, $[n] = \{1, 2, \ldots, n\}$. The formula $d \,|\, n$ denotes the relation "d divides n," i. e., $(\exists k)(n = dk)$. We also say "d is a divisor of n" or "n is a multiple of d." Note that $(\forall a)(a \,|\, a)$, including $0 \,|\, 0$ (even though we do not allow division by zero!). In fact $0 \,|\, n \iff n = 0$. Note also that $(\forall k \ (n \,|\, k)) \iff n = \pm 1$.

Notation 4.1.1. Let $\operatorname{div}(n)$ denote the set of divisors of n.

Examples. $\operatorname{div}(6) = \operatorname{div}(-6) = \{\pm 1, \pm 2, \pm 3, \pm 6\}$; $\operatorname{div}(1) = \{\pm 1\}$; $\operatorname{div}(0) = \mathbb{Z}$.

Exercise 4.1.2. Prove: $a \mid b \quad \iff \quad \operatorname{div}(a) \subseteq \operatorname{div}(b)$.

Exercise 4.1.3. Prove: $\operatorname{div}(a) = \operatorname{div}(b) \quad \iff \quad b = \pm a$.

Congruence notation. We write $a \equiv b \pmod{m}$ if $m \,|\, (a - b)$ ("a is congruent to b modulo m").

For instance, $100 \equiv 2 \pmod{7}$ (because $7 \mid 100 - 2 = 98 = 7 \cdot 14$); therefore, if today is Monday then 100 days from now it will be Wednesday (Monday $+2$). This example expains why *modular arithemtic* (calculations modulo m) are also referred to as "calendar arithmetic."

Division Theorem. $(\forall a)(\forall b \geq 1)(\exists q)(\exists r)(0 \leq r \leq b - 1 \text{ and } a = bq + r)$.

q is called the "integer quotient" and r the "remainder."

Exercise 4.1.4. Prove: $r \equiv a \pmod{b}$.

Remainder notation. The remainder r is denoted by the expression $(a \mod b)$. (Exercise 4.1.4 explains this notation; the congruence *relation* and the mod *function* should not be confused.) Examples: $(100 \mod 7) = 2$; $(-100 \mod 7) = 5$; $(98 \mod 7) = 0$; $(0 \mod 7) = 0$; $(a \mod 0)$ is undefined.

Common Divisor. The integer f is a common divisor of the integers a and b if $f \mid a$ and $f \mid b$.

Exercise 4.1.5. Prove: f is a common divisor of a and $b \iff \operatorname{div}(f) \subseteq \operatorname{div}(a) \cap \operatorname{div}(b)$.

Greatest Common Divisor. The integer d is a greatest common divisor of the integers a and b if

- d is a common divisor of a and b;
- every common divisor of a and b divides d.

Exercise 4.1.6. Prove: d is a greatest common divisor of a and $b \iff \operatorname{div}(d) = \operatorname{div}(a) \cap \operatorname{div}(b)$.

The existence of a greatest common divisor is not evident at all; it is an important basic theorem. Often we need the additional fact that the greatest common divisor can be written as a linear combination with integer coefficients: $d = au + bv$.

Exercise$^+$ 4.1.7. $(\forall a)(\forall b)(\exists u)(\exists v)(au + bv$ is a greatest common divisor of a and $b)$.

Exercise 4.1.8. Prove: if d is a greatest common divisor of a and b then $-d$ is also a greatest common divisor of a and b and there are no other greatest common divisors.

G.c.d. notation. g.c.d.(a, b) will denote the (unique) nonnegative greatest common divisor of the integers a and b.

Exercise 4.1.9. Prove: g.c.d.$(0, 0) = 0$.

Exercise 4.1.10. What are the common divisors of 0 and 0? Is 0 the "greatest"?

Exercise 4.1.11. (a) Prove: $(\forall a)($g.c.d.$(a, a) = |a|)$.

(b) Prove: $(\forall a)($g.c.d.$(a, 0) = |a|)$.

Note that each of these statements includes the fact that g.c.d.$(0, 0) = 0$.

The **Euclidean algorithm**, described in Euclid's *Elements* around 350 B.C.E., is an efficient method to calculate the g.c.d. of two positive integers. We describe the algorithm in *pseudocode.*

Euclidean Algorithm

INPUT: integers a, b.

OUTPUT: g.c.d.(a, b).

```
Initialize: $A := |a|$, $B := |b|$
    while $B \geq 1$ do
        division: $R := (A \mod B)$
        $A := B$, $B := R$
    end(while)
return $A$
```

The **correctness** of the algorithm follows from the following *loop invariant:*

$$\text{g.c.d.}(A, B) = \text{g.c.d.}(a, b).$$

Exercise 4.1.12. Prove that the statement above is indeed a *loop invariant*, i. e., prove that if the statement "g.c.d.(A, B) =g.c.d.(a, b)" is true before an iteration of the **while** loop then it remains true after the execution of the **while** loop.

In addition, at the end we use the fact that g.c.d.$(A, 0) = A$.

Exercise 4.1.13. The **efficiency** of the Euclidean the algorithm follows from the observation that after every two rounds, the value of B is reduced to less than half. Prove this statement.

This implies that the number of rounds is $\leq 2n$ where n is the number of binary digits of b. Therefore the total number of bit-operations is $O(n^3)$, so this is a *polynomial-time algorithm.* (Good job, Euclid!)

Exercise 4.1.14. Use Euclid's algorithm to determine the g.c.d. of the following pairs of integers:

(a) (105; 480)

(b) (72,806; 13,587,574).

Exercise 4.1.15. Let n be a *positive* integer and let $d(n)$ denote the number of positive divisors of n. For instance, $d(1) = 1$, $d(2) = d(3) = d(5) = 2$, $d(4) = 3$, $d(6) = 4$. Prove your answers to the following questions.

(a) For what values of n is $d(n) = 2$?

(b) For what values of n is $d(n) = 3$?

(c) Prove: $(\forall n)(d(n) < 2\sqrt{n})$.

Exercise 4.1.16. (a) Let $a, b > 0$ and let us perform Euclid's algorithm to find the g.c.d. of a and b. Let $r_1, r_2, \ldots$ denote the successive remainders; let us use the notation $r_{-1} = a$ and $r_0 = b$. Prove: $(\forall i \geq -1)(r_{i+2} \leq r_i/2)$.

(b) Prove: if a has n bits (digits in binary) then the algorithm will terminate in $\le 2n$ rounds (one round being a division to find the next remainder). *Hint:* use part (a).

Exercise 4.1.17. Recall that the *multiplicative inverse* of b modulo m, denoted by $x = (b^{-1} \pmod{m})$, is an integer x such that $bx \equiv 1 \pmod{m}$. Find each of the following multiplicative inverses, or prove that the multiplicative inverse does not exist. Among the infinitely many values of the multiplicative inverse, find the smallest positive integer.

(a) $5^{-1} \pmod{17}$

(b) $39^{-1} \pmod{403}$

(c) $2^{-1} \pmod{2k+1}$ (where k is a given integer).

(d) $k^{-1} \pmod{2k+1}$. Find the inverse in the range $\{0, 1, \dots, 2k\}$.

(e) $k^{-1} \pmod{3k+1}$. Find the inverse in the range $\{0, 1, \dots, 3k\}$.

Exercise 4.1.18. Solve the following system of congruences:

$$\begin{aligned} x &\equiv 7 \pmod{16} \\ x &\equiv 3 \pmod{15} \\ x &\equiv 1 \pmod{11} \end{aligned}$$

Exercise 4.1.19. Decide whether or not the following system of congruences is solvable. If your answer is YES, find a solution. If your answer is NO, prove your answer.

$$\begin{aligned} x &\equiv 7 \pmod{13} \\ x &\equiv 3 \pmod{25} \\ x &\equiv 20 \pmod{39} \end{aligned}$$

Exercise 4.1.20. Prove whether or not the following system of congruences is solvable.

$$\begin{aligned} x &\equiv 7 \pmod{18} \\ x &\equiv 7 \pmod{12} \\ x &\equiv 1 \pmod{6} \end{aligned}$$

Exercise 4.1.21. Consider the statement "if $a \equiv 1 \pmod{5}$ and $b \equiv 1 \pmod{5}$ then g.c.d.$(a, b) \equiv 1 \pmod{5}$." Find infinitely many counterexamples.

Exercise 4.1.22. The **Fibonacci numbers** are defined as follows: $F_0 = 0, F_1 = 1$, and for $n \ge 2$, $F_n = F_{n-1} + F_{n-2}$. So $F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5, F_6 = 8$, $F_7 = 13$, $F_8 = 21$, etc. Prove: for all $n \ge 1$,

(a) $\gcd(F_{n-1}, F_n) = 1$.

(b) $|F_n^2 - F_{n-1}F_{n+1}| = 1$.

(c) If $\gcd(m,n) = d$ then $\gcd(F_m, F_n) = F_d$.

(d) If $\gcd(m,n) = d$ then $\gcd(a^m - 1, a^n - 1) = a^d - 1$.

Hint: For parts (a) and (b), use mathematical induction.

Exercise 4.1.23. Calculate ($a \bmod m$) where $a = 3^{114{,}555}$ and $m = 173$. Recall that the expression (a mod m) denotes the smallest nonnegative remainder of the division of a by m.
Hint. Fermat's little Theorem (Theorem 4.2.18).

Exercise 4.1.24. (a) Prove: if m is a prime and $x^2 \equiv 1 \pmod{m}$ then $x \equiv \pm 1 \pmod{m}$ (i. e., either $x \equiv 1 \pmod{m}$, or $x \equiv -1 \pmod{m}$).

(b) Prove that (a) becomes false if we omit the condition that m is a prime. (Give a counterexample.)

(c) Prove that (a) is false for every m of the form $m = pq$ where p, q are distinct odd primes. In other words, show that $(\forall p,q)(\exists x)($ if p, q are distinct odd primes then $x^2 \equiv 1 \pmod{pq}$ but $x \not\equiv \pm 1 \pmod{pq})$). *Hint.* Observe that $a \equiv b \pmod{pq} \Leftrightarrow a \equiv b \pmod{p}$ and $a \equiv b \pmod{q}$. Work separately modulo each prime; combine your results using the Chinese Remainder Theorem.

Exercise 4.1.25. Prove: $\forall x (x^2 \not\equiv -1 \pmod{419})$.
Hint. Proof by contradiction. Use Fermat's little Theorem (Theorem 4.2.18). (419 is a prime.)

Exercise 4.1.26. (a) Prove: if g.c.d.$(a, 85) = 1$ then $a^{33} \equiv a \pmod{85}$). *Hint.* $85 = 5 \cdot 17$, so two numbers are congruent modulo 85 if and only if they are congruent modulo 5 as well as modulo 17. Prove the stated congruence modulo 5 and modulo 17.

(b) True or false (prove your answer): if 85 does not divide a then $a^{32} \equiv 1 \pmod{85}$).

Exercise 4.1.27. True or False. If False, give a counterexample.

1. If $\gcd(a, b) = 0$ then $a = b = 0$.
2. If l.c.m. $(a, b) = 0$ then $a = b = 0$.
3. If $a \equiv b \pmod{24}$ then $a \equiv b \pmod{6}$ and $a \equiv b \pmod{4}$.
4. If $a \equiv b \pmod{6}$ and $a \equiv b \pmod{4}$ then $a \equiv b \pmod{24}$.

Exercise 4.1.28. Consider the following statement:

Statement. a^{15} is a multiplicative inverse of a modulo 17.

1. Define what it means that "x is a multiplicative inverse of a modulo m."

2. Give infinitely many counterexamples to the statement above.
3. State a very simple necessary and sufficient condition for the statement to be true. Prove your answer.

Exercise 4.1.29. Prove: $(\forall a)(a^{37} \equiv a \pmod{247})$. *Hint.* $247 = 13 \cdot 19$.

Exercise 4.1.30. Prove: if a is an odd integer then

$$a^{67} \equiv a \pmod{12,328}.$$

Hint. $12,328 = 8 \cdot 23 \cdot 67$.

Exercise 4.1.31. Prove: the congruence $x^2 \equiv -1 \pmod{103}$ has no solution. (103 is a prime number.) *Hint.* FℓT.

Exercise 4.1.32. Let $1 \le a_1 < \cdots < a_{n+1} \le 2n$ be $n+1$ distinct integers between 1 and $2n$. Prove:

(a) $(\exists i, j)(i \neq j$ and g.c.d.$(a_i, a_j) = 1)$.

(b) $(\exists i, j)(i \neq j$ and $a_i \mid a_j)$. *Hint.* Pigeon-hole principle.

Exercise 4.1.33. Let p be a prime number. Find all solutions to the following congruence. Prove your answer.

$$x^p \equiv x^{3p} \pmod{p}.$$

Exercise 4.1.34. In this problem, the universe of the variable x is the set of integers. Prove:

$$(\forall x)(x^{21} \equiv x \pmod{55}).$$

4.2 Gcd, congruences

Exercise 4.2.1. Prove that the product of n consecutive integers is always divisible by $n!$. *Hint.* One-line proof.

Exercise 4.2.2. (The Divisor Game) Select an integer $n \ge 2$. Two players alternate naming positive divisors of n subject to the following rule: no divisor of any previously named integer can be named. The first player forced to name "n" loses. Example: if $n = 30$ then the following is a possible sequence of moves: 10, 3, 6, 15, at which point it is the first player's move; he is forced to say "30" and loses.

1. Find a winning strategy for the first player when n is a prime power; or of the form pq^k; p^kq^k; pqr; or $pqrs$, where p, q, r, s are prime and k is a positive integer.

2. Prove: $\forall n \geq 2$, the first player has a winning strategy. (*Hint:* prove, in two or three lines, the *existence* of a winning strategy.)

Notation 4.2.3. Let $\mathrm{Div}\,(n)$ denote the set of positive divisors of n.

Exercise 4.2.4. Prove, for all $a, b \in \mathbb{Z}$,

$$(\mathrm{Div}\,(a) \subseteq \mathrm{Div}\,(b)) \quad \Longleftrightarrow \quad a \mid b.$$

Exercise$^+$ 4.2.5. Prove: $(\forall a, b)(\exists d)(\mathrm{Div}\,(a) \cap \mathrm{Div}\,(b) = \mathrm{Div}\,(d))$. A nonnegative d satisfying this statement is called the g.c.d. of a and b. Note that $\gcd(a, b) = 0 \Longleftrightarrow a = b = 0$. Define l.c.m. analogously. When is $\mathrm{l.c.m.}\,(a, b) = 0$?

Exercise 4.2.6. Prove: $\gcd(a^k - 1, a^\ell - 1) = a^d - 1$, where $d = \gcd(k, \ell)$.

Definition 4.2.7. The Fibonacci numbers are defined by the recurrence $F_n = F_{n-1} + F_{n-2}$, $F_0 = 0$, $F_1 = 1$.

Exercise$^+$ 4.2.8. Prove: $\gcd(F_k, F_\ell) = F_d$, where $d = \gcd(k, \ell)$.

Exercise 4.2.9. Prove: if $a \equiv b \pmod{m}$ then $\gcd(a, m) = \gcd(b, m)$.

Exercise 4.2.10. Prove: if $a, b \geq 0$ then $\gcd(a, b) \cdot \mathrm{l.c.m.}\,(a, b) = ab$.

Exercise 4.2.11. Prove: congruence modulo m is an equivalence relation on $\mathbb{Z}$. The equivalence classes are called the *residue classes* mod m. There are m residue classes modulo m. Under the natural operations they form the ring $\mathbb{Z}/m\mathbb{Z}$. The additive group of this ring is cyclic.

Exercise 4.2.12. Prove that the sequence of Fibonacci numbers mod m is periodic. The length of the period is $\leq m^2 - 1$.

Exercise 4.2.13. An *integer-preserving polynomial* is a polynomial $f(x)$ such that $(\forall a \in \mathbb{Z})(f(a) \in \mathbb{Z})$. Prove that $f(x)$ is integer-preserving if and only if it can be written as

$$f(x) = \sum_{i=0}^{n} a_i \binom{x}{i} \tag{4.1}$$

with suitable integer coefficients a_i. Here

$$\binom{x}{i} = \frac{x(x-1)\dots(x-i+1)}{i!}; \qquad \binom{x}{0} = 1.$$

Exercise 4.2.14. A *congruence-preserving polynomial* is an integer-preserving polynomial such that $(\forall a, b, m \in \mathbb{Z})(a \equiv b \pmod{m} \Rightarrow f(a) \equiv f(b) \pmod{m})$. Prove that $f(x)$ is congruence-preserving if and only if $(\forall i)(e_i \mid a_i)$ in the expression (4.1), where $e_i = \mathrm{l.c.m.}\,(1, 2, \dots, i)$.

Exercise 4.2.15. A *multiplicative inverse* of a modulo m is an integer x such that $ax \equiv 1 \pmod{m}$; notation: $x = a^{-1} \bmod m$. Prove: $\exists a^{-1} \bmod m \Longleftrightarrow \gcd(a, m) = 1$.

Exercise 4.2.16. (Wilson's theorem) Prove: $(p-1)! \equiv -1 \pmod{p}$. *Hint:* match each number with its multiplicative inverse in the product $(p-1)!$

Exercise 4.2.17. Prove: if $\gcd(a,p)=1$ then $\prod_{j=1}^{p-1} j \equiv \prod_{i=1}^{p-1}(ai)$. *Hint.* Match terms on the right hand side with terms on the left hand side so that corresponding terms satisfy $j \equiv ai \pmod{p}$.

Theorem 4.2.18 (Fermat's little Theorem)**.** If $\gcd(a,p)=1$ then $a^{p-1} \equiv 1 \pmod{p}$.

Exercise 4.2.19. Infer Fermat's little Theorem from Exercise 4.2.17.

Exercise 4.2.20. Use the same idea to prove the **Euler–Fermat theorem:** if $\gcd(a,m)=1$ then $a^{\varphi(m)} \equiv 1 \pmod{m}$. ($\varphi$ is Euler's φ function, see Definition 4.3.1).

Exercise 4.2.21. Prove: if p is a prime and f is a polynomial with integer coefficients then $f(x)^p \equiv f(x^p) \pmod{p}$. Here the congruence of two polynomials means coefficientwise congruence.

The multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$ consists of the mod m residue classes relatively prime to m. Its order is $\varphi(m)$. For a review of related concepts in abstract algebra, see Chapter **??** (cf. especially Exercise **??**).

Exercise$^+$ 4.2.22. Prove: if p is a prime then $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic (see Definition **??**). A generator of this group is called a *primitive root mod p*.

Exercise$^+$ 4.2.23. Prove: if p is an odd prime then $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is cyclic.

Exercise$^+$ 4.2.24. If $k \geq 2$ then the group $(\mathbb{Z}/2^k\mathbb{Z})^\times$ is not cyclic but the direct sum of a cyclic group of order 2 and a cyclic group of order 2^{k-2}.

4.3 Arithmetic Functions

Definition 4.3.1 (Euler's Phi Function)**.**

$$\begin{aligned} \varphi(n) &= \left|\left\{k \in [n] \,:\, \gcd(k,n)=1\right\}\right| \\ &= \text{number of positive integers not greater than } n \text{ which are relatively prime to } n \end{aligned}$$

Exercise 4.3.2. Show that the number of complex primitive n-th roots of unity is $\varphi(n)$. Show that if $d|n$ then the number of elements of order d in a cyclic group of order n is $\varphi(d)$.

Exercise 4.3.3. Show

$$\sum_{d\,|\,n} \varphi(d) = n.$$

Exercise$^+$ 4.3.4. Let $D_n = (d_{ij})$ denote the $n \times n$ matrix with $d_{ij} = \gcd(i, j)$. Prove:

$$\det D_n = \varphi(1)\varphi(2) \cdots \varphi(n).$$

(*Hint.* Let $Z = (z_{ij})$ be the matrix with $z_{ij} = 1$ if $i \mid j$ and $z_{ij} = 0$ otherwise. Consider the matrix $Z^T FZ$ where F is the diagonal matrix with entries $\varphi(1), \dots, \varphi(n)$ and Z^T is "Z-transpose" (reflection in the main diagonal).)

Definition 4.3.5 (Number of [positive] divisors)**.**

$$d(n) = \Big|\{d \in \mathbb{N} : d \mid n\}\Big|$$

Exercise 4.3.6. Prove: $d(n) < 2\sqrt{n}$.

Exercise$^+$ 4.3.7. Prove: $(\forall \epsilon > 0)(\exists n_0)(\forall n > n_0)(d(n) < n^\epsilon)$. (*Hint.* Use a consequence of the Prime Number Theorem (Theorem 4.4.6 in the next section).) Prove that $d(n) < n^{c/\ln\ln n}$ for some constant c. The best asymptotic constant is $c = \ln 2 + o(1)$.

Exercise$^+$ 4.3.8. Prove that for infinitely many values of n the reverse inequality $d(n) > n^{c/\ln\ln n}$ holds (with another constant $c > 0$). (Again, use the PNT.)

Exercise$^+$ 4.3.9. Let $D(n) = (1/n)\sum_{i=1}^n d(i)$ (the average number of divisors). Prove: $D(n) \sim \ln(n)$. (*Comment.* If we pick an integer t at random between 1 and n then $D(n)$ will be the *expected number* of divisors of t. – Make your proof very simple (3 lines). Do not use the PNT.)

Exercise$^+$ 4.3.10. Prove: $(1/n)\sum_{i=1}^n d(i)^2 = \Theta((\ln n)^3)$.

Definition 4.3.11 (Sum of [positive] divisors)**.**

$$\sigma(n) = \sum_{d \mid n} d$$

Definition 4.3.12 (Number of [distinct] prime divisors)**.** Let $n = p_1^{k_1} \cdots p_r^{k_r}$ where the p_i are distinct primes and $k_i > 0$. Set $\nu(n) = r$ (number of distinct prime divisors; so $\nu(1) = 0$). Set $\nu^*(n) = k_1 + \cdots + k_r$ (total number of prime divisors; so $\nu^*(1) = 0$).

Exercise$^+$ 4.3.13. Prove that the expected number of distinct prime divisors of a random integer $i \in [n]$ is asymptotically $\ln\ln n$:

$$\frac{1}{n}\sum_{i=1}^{n} \nu(i) \sim \ln\ln n.$$

How much larger is ν^*? On average, not much. Prove that the average value of ν^* is also asymptotic to $\ln\ln n$.

Definition 4.3.14. n is **square-free** if $(\forall p \text{ prime })(p^2 \nmid n)$.

Definition 4.3.15 (Möbius Function)**.**

$$\mu(n) = \begin{cases} 1 & n = 1 \\ (-1)^k & n = p_1 \cdots p_k \text{ where the } p_i \text{ are distinct } (n \text{ is square-free}) \\ 0 & \text{if } (\exists p)(p^2 \mid n) \end{cases}$$

Exercise 4.3.16. Let $\delta(n) = \sum_{d|n} \mu(d)$. Evaluate $\delta(n)$.

Definition 4.3.17 (Riemann zeta function)**.** For $s > 1$ define the *zeta function* $\zeta(s) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s}$.

Exercise 4.3.18. Prove Euler's identity:

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1 - \frac{1}{p^s}}.$$

Exercise 4.3.19. Prove:

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}.$$

Exercise 4.3.20. Prove:

$$(\zeta(s))^2 = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}.$$

Exercise 4.3.21. Prove:

$$\zeta(s)(\zeta(s) - 1) = \sum_{n=1}^{\infty} \frac{\sigma(n)}{n^s}.$$

Exercise* 4.3.22. (Euler) Prove: $\zeta(2) = \pi^2/6$.

Exercise 4.3.23. Give a natural definition which will make following statement sensible and true: "the probability that a random positive integer n satisfies $n \equiv 3 \pmod 7$ is 1/7." Our choice of a "random positive integer" should be "uniform" (obviously impossible). (*Hint.* Consider the integers up to x; then take the limit as $x \to \infty$.)

Exercise 4.3.24. Make sense out of the question "What is the probability that two random positive integers are relatively prime?" Prove that the answer is $6/\pi^2$. *Hint.* To prove that the required limit exists may be somewhat tedious. If you want to see the fun part, assume the existence of the limit, and prove in just two lines that the limit must be $1/\zeta(2)$.

Definition 4.3.25. Let F be a field. $f : \mathbb{N} \to F$ is called **multiplicative** if

$$(\forall a, b)(\gcd(a, b) = 1 \Rightarrow f(ab) = f(a)f(b)).$$

f is called **completely multiplicative** if

$$(\forall a, b)(f(ab) = f(a)f(b)).$$

f is called **additive** if

$$(\forall a, b)(\gcd(a, b) = 1 \Rightarrow f(ab) = f(a) + f(b)).$$

Exercise 4.3.26. Show that

1. φ, σ, d, and μ are multiplicative but not completely multiplicative
2. ν is additive and ν^* is completely additive. Log is completely additive.

Exercise 4.3.27. Show

1. $\varphi\left(p^k\right) = p^k - p^{k-1} = (p-1)p^{k-1}$
2. $d\left(p^k\right) = k+1$
3. $\sigma\left(p^k\right) = \dfrac{p^{k+1}-1}{p-1}$

Exercise 4.3.28. Show

1. $\varphi\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r}(p_i-1)p_i^{k_i-1}$
2. $d\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r}(k_i+1)$
3. $\sigma\left(\prod_{i=1}^{r} p_i^{k_i}\right) = \prod_{i=1}^{r}\dfrac{p_i^{k_i+1}-1}{p_i-1}$

Exercise 4.3.29. Show

$$\varphi(n) = n \prod_{\substack{p \mid N \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right)$$

Let F be a field and $f\colon \mathbb{N} \to F$. Define

$$g(n) = \sum_{d \mid n} f(d).$$

Exercise 4.3.30 (Möbius Inversion Formula)**.** Show

$$f(n) = \sum_{d \mid N} g(d)\mu\left(\frac{n}{d}\right).$$

Exercise 4.3.31. Use the Möbius Inversion Formula together with Exercise 4.3.3 for a second proof of Exercise 4.3.29.

Exercise 4.3.32. Prove that the sum of the complex primitive n-th roots of unity is $\mu(n)$.

Definition 4.3.33. The n**-th cyclotomic polynomial** $\Phi_n(x)$ is defined as

$$\Phi_n(x) = \prod_{\omega}(x - \omega)$$

where the product ranges over all complex primitive n-th roots of unity. Note that the degree of $\Phi_n(x)$ is $\varphi(n)$. Also note that $\Phi_1(x) = x-1$, $\Phi_2(x) = x+1$, $\Phi_3(x) = x^2+x+1$, $\Phi_4(x) = x^2+1$, $\Phi_5(x) = x^4+x^3+x^2+x+1$, $\Phi_6(x) = x^2-x+1$.

Exercise 4.3.34. Prove that $\Phi_n(x)$ has integer coefficients. What is the coefficient of $x^{\varphi(n)-1}$?

Exercise 4.3.35. Prove: if p is a prime then $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$.

Exercise 4.3.36. Prove:

$$\Phi_n(x) = \prod_{d|n}(x^d - 1)^{\mu(n/d)}.$$

Exercise$^+$ 4.3.37. (Bateman) Let A_n denote the sum of the absolute values of the coefficients of $\Phi_n(x)$. Prove that $A_n < n^{d(n)/2}$. Infer from this that $A_n < \exp(n^{c/\ln\ln n})$ for some constant c. *Hint:* We say that the power series $\sum_{n=0}^{\infty} a_n x^n$ *dominates* the power series $\sum_{n=0}^{\infty} b_n x^n$ if $(\forall n)(|b_n| \le a_n)$. Prove that the power series

$$\prod_{d|n} \frac{1}{1-x^d}$$

dominates $\Phi_n(x)$.

Note: Erdős proved that this bound is tight, apart from the value of the constant: for infinitely many values of n, $A_n > \exp(n^{c/\ln\ln n})$ for another constant $c > 0$.

Exercise$^+$ 4.3.38. (Hermite) Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a monic polynomial of degree n (i. e., $a_n = 1$) with integer coefficients. Suppose all roots of f have unit absolute value. Prove that all roots of f are roots of unity. (In other words, if all algebraic conjugates of a complex algebraic number z have unit absolute value then z is a root of unity.)

4.4 Prime Numbers

Exercise 4.4.1. Prove:

$$\sum_{i=1}^{n} \frac{1}{i} = \ln n + O(1).$$

Exercise 4.4.2. Prove:

$$\prod_{p \le x} \frac{1}{1 - 1/p} = \sum{}' \frac{1}{i},$$

where the product is over all primes $\le x$ and the summation extends over all positive integers i with no prime divisors greater than x. In particular, the sum on the right-hand side converges. It also follows that the left-hand side is greater than $\ln x$.

Exercise 4.4.3. Prove: $\sum 1/p = \infty$. (*Hint.* Use the preceding exercise. Take natural logarithms; use the power series expansion of $\ln(1-z)$. Conclude that $\sum_{p\le x} 1/p > \ln\ln x + O(1)$. (In other words, $\sum_{p\le x} 1/p - \ln\ln x$ is bounded from below.))

Exercise$^+$ 4.4.4. Prove: $\sum_{p\le x} 1/p = \ln\ln x + O(1)$. (In other words, $|\sum_{p\le x} 1/p - \ln\ln x|$ is bounded.)

Exercise$^+$ 4.4.5. Prove $\varphi(n) = \Omega\left(\frac{n}{\ln\ln n}\right)$ and find the largest implicit asymptotic constant.

Let $\pi(x)$ the number of primes less than or equal to x.

Theorem 4.4.6 (Prime Number Theorem)(Hadamard and de la Vallée Poussin, 1896)**.**

$$\pi(x) \sim \frac{x}{\ln x}$$

Exercise 4.4.7. Use the PNT to show that $\lim\limits_{n\to\infty} \frac{p_{n+1}}{p_n} = 1$, where p_n is the n-th prime.

Exercise 4.4.8. Use the PNT to prove $p_n \sim n \cdot \ln n$.

Exercise 4.4.9. Prove $\prod\limits_{\substack{p\le x\\ p \text{ prime}}} p = \exp\big(x(1+o(1))\big)$. Prove that this result is in fact equivalent to the PNT.

Exercise 4.4.10. Let $e_n = \text{l.c.m.}\,(1, 2, \dots, n)$. Prove: $e_n = \exp\big(n(1+o(1))\big)$. Prove that this result is in fact equivalent to the PNT.

Exercise 4.4.11. Prove: $\sum_{p\le x} p \sim x^2/(2\ln x)$. (Use the PNT.)

Definition 4.4.12. A *permutation* is a bijection of a set to itself. The permutations of a set form a group under composition. The *symmetric group of degree n* is the group of all permutations of a set of n elements; it has order $n!$. The *exponent* of a group is the l.c.m. of the orders of all elements of the group.

Exercise 4.4.13. Prove: the exponent of S_n is e_n.

Exercise$^+$ 4.4.14. Let $m(n)$ denote the maximum of the orders of the elements in S_n. Prove: $m(n) = \exp(\sqrt{n\ln n}(1+o(1)))$.

Exercise* 4.4.15. Let $a(n)$ denote the "typical" order of elements in S_n. Prove that $\ln a(n) = O((\ln n)^2)$. ("Typical" order means that 99% of the elements has order falling in the stated range. Here "99" is arbitrarily close to 100.) *Hint.* Prove that a typical permutation has $O(\ln n)$ cycles.

Erdős and Turán proved in 1965 that in fact $\ln a(n) \sim (\ln n)^2/2$.

Exercise 4.4.16. Prove from first principles: $\prod_{\substack{p<x \\ p \text{ prime}}} p < 4^x$. (*Hint:* if $n < p \le 2n$ then $p \mid \binom{2n}{n}$.)

Exercise 4.4.17. Prove: if $p > \sqrt{2n}$ then $p^2 \nmid \binom{2n}{n}$.

Exercise 4.4.18. Prove: if q is a prime power dividing $\binom{2n}{n}$ then $q \le n$. (*Hint.* Give a formula for the highest exponent of a prime p which divides $\binom{2n}{n}$. First, find a formula for the exponent of p in $n!$.)

Exercise 4.4.19. Prove from first principles: $\prod_{\substack{p<x \\ p \text{ prime}}} p > (2 + o(1))^x$. (*Hint.* Consider the prime-power decomposition of $\binom{x}{x/2}$. Show that the contribution of the powers of primes $\le \sqrt{x}$ is negligible.)

Exercise 4.4.20. Paul Erdős was an undergraduate when he found a simple proof of Chebyshev's theorem based on the prime factors of $\binom{2n}{n}$. Chebyshev's theorem is a precursor of the PNT; it says that

$$\pi(x) = \Theta\left(\frac{x}{\ln x}\right).$$

Following Erdős, prove Chebyshev's Theorem from first principles. The proof should be only a few lines, based on Exercises 4.4.16 and 4.4.19.

Exercise 4.4.21. Prove: for all integers x, either $x^2 \equiv 0 \pmod 4$ or $x^2 \equiv 1 \pmod 4$. (*Hint.* Distinguish two cases according to the parity of x [parity: even or odd].)

Exercise 4.4.22. $a^2 + b^2 \not\equiv -1 \pmod 4$.

Exercise 4.4.23. (a) Make a table of all primes ≤ 100. Next to each prime p write its expression as the sum of two squares if p can be so represented; otherwise write "NONE" next to p.

(b) Discover and state a very simple pattern as to which primes can and which primes cannot be represented as the sum of two squares. Your statement should go like this: "It seems from the table that a prime p can be represented as the sum of two squares if and only if either $p = 2$ or ***" where "***" stands for a very simple rule (less than half a line).

(c) Give a simple proof that the primes you believe cannot be represented as a sum of two squares indeed cannot. *Hint.* Use the previous exercise.

Exercise 4.4.24. Prove: if p is a prime number and $p \ge 5$ then $p \equiv \pm 1 \pmod 6$. *Hint.* There are only 6 cases to consider. (What are they?)

4.5 Quadratic Residues

Definition 4.5.1. a is a **quadratic residue** mod p if $(p \nmid a)$ and $(\exists b)(a \equiv b^2 \bmod p)$.

Exercise 4.5.2. Prove: a is a quadratic residue mod $p \iff a^{(p-1)/2} \equiv 1 \pmod{p}$.

Definition 4.5.3. a is a **quadratic non-residue** mod p if $(\forall b)(a \not\equiv b^2 \bmod p)$.

Exercise 4.5.4. Prove: a is a quadratic non-residue mod $p \iff a^{(p-1)/2} \equiv -1 \pmod{p}$.

Definition 4.5.5 (Legendre Symbol)**.**

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue mod } p \\ -1 & \text{if } a \text{ is a quadratic non-residue mod } p \\ 0 & \text{if } p \mid a \end{cases}$$

Let $\mathbb{F}_q$ be a finite field of odd prime power order q.

Definition 4.5.6. $a \in F_q$ is a **quadratic residue** if $a \neq 0$ and $(\exists b)(a = b^2)$.

Exercise 4.5.7. Prove: a is a quadratic residue in $\mathbb{F}_q \iff a^{(q-1)/2} = 1$.

Definition 4.5.8. $a \in \mathbb{F}_q$ is a **quadratic non-residue** if $(\forall b)(a \neq b^2)$.

Exercise 4.5.9. Prove: a is a quadratic non-residue in $\mathbb{F}_q \iff a^{(q-1)/2} = -1$.

Exercise 4.5.10. Prove: in $\mathbb{F}_q$, the number of quadratic residues equals the number of quadratic non-residues; so there are $(q-1)/2$ of each. (As before, q is an odd prime power.)

Definition 4.5.11. Let q be an odd prime power. We define the **quadratic character** $\chi\colon \mathbb{F}_q \to \{0, 1, -1\} \subset \mathbb{C}$ by

$$\chi(a) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue} \\ -1 & \text{if } a \text{ is a non-residue} \\ 0 & \text{if } a = 0 \end{cases}$$

Note that if $q = p$ (i.e. prime and not prime power) then $\chi(a) = \left(\frac{a}{p}\right)$.

Exercise 4.5.12. Prove χ is multiplicative.

Exercise 4.5.13. The Legendre Symbol is completely multiplicative in the numerator.

Exercise 4.5.14. Prove that -1 is a quadratic residue in $\mathbb{F}_q$ if and only if $q \equiv 1 \pmod{4}$. *Hint.* Exercise 4.5.7.

Exercise 4.5.15. Prove that $\sum_{a \in \mathbb{F}_q} \chi(a(a-1)) = -1$. *Hint.* Divide by a^2.

Exercise 4.5.16. Prove that each of the four pairs $(\pm 1, \pm 1)$ occur a roughly equal number of times ($\approx q/4$) as $(\chi(a), \chi(a-1))$ $(a \in \mathbb{F}_q)$. "Roughly equal" means the difference is bounded by a small constant. Moral: for a random element $a \in \mathbb{F}_q$, the values of $\chi(a)$ and $\chi(a-1)$ are nearly independent.

Exercise 4.5.17. Let $f(x) = ax^2 + bx + c$ be a quadratic polynomial over $\mathbb{F}_q$ $(a, b, c \in \mathbb{F}_q$, $a \neq 0)$. Prove: if $b^2 - 4ac \neq 0$ then $|\sum_{a \in \mathbb{F}_q} \chi(f(a))| \leq 2$. What happens if $b^2 - 4ac = 0$?

4.6 Lattices and diophantine approximation

Definition 4.6.1. An n-dimensional **lattice** (grid) is the set L of all *integer* linear combinations $\sum_{i=1}^{n} a_i \mathbf{b}_i$ of a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of $\mathbb{R}^n$ ($a_i \in \mathbb{Z}$). The set of *real* linear combinations with $0 \le a_i \le 1$ ($a_i \in \mathbb{R}$) form a **fundamental parallelepiped**.

Exercise 4.6.2. The volume of the fundamental parallelepiped of the lattice L is $\det(L) := |\det(\mathbf{b}_1, \dots, \mathbf{b}_n)|$.

Exercise* 4.6.3. (Minkowski's Theorem) Let L be an n-dimensional lattice and let V be the volume of its fundamental parallelepiped. Let $A \subset \mathbb{R}^n$ be an n-dimensional convex set, symmetrical about the origin (i. e., $-A = A$), with volume greater than $2^n V$. Then $A \cap L \neq \{0\}$, i. e., A contains a lattice point other than the origin.
Hint. Linear transformations don't change the proportion of volumes, and preserve convexity and central symmetry. So WLOG $L = \mathbb{Z}^n$ with $\{\mathbf{b}_i\}$ the standard basis. The fundamental parallelepiped is now the unit cube C. Consider the lattice $2L = (2\mathbb{Z})^n$. Then the quotient space $\mathbb{R}^n/(2\mathbb{Z})^n$ can be identified with the cube $2C$ which has volume 2^n. Since A has volume $> 2^n$, there exist two points $u, v \in A$ which are mapped to the same point in $2C$, i. e., all coordinates of $u - v$ are even integers. Show that $(u - v)/2 \in A \cap L$.

Exercise 4.6.4. Finding "short" vectors in a lattice is of particular importance. Prove the following corollary to Minkowski's Theorem:

$$(\exists v \in L)\left(0 < \|v\|_\infty \le (\det L)^{1/n}\right).$$

Definition 4.6.5. Let $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. A *simultaneous ϵ-approximation* of the α_i is a sequence of fractions p_i/q with a common denominator $q > 0$ such that $(\forall i)(|q\alpha_i - p_i| \le \epsilon)$.

Exercise$^+$ 4.6.6. (Dirichlet) $(\forall \alpha_1, \dots, \alpha_n \in \mathbb{R})(\forall \epsilon > 0)(\exists$ an ϵ-approximation with the denominator satisfying $0 < q \le \epsilon^{-n})$.
Hint. Apply the preceding exercise to the $(n+1)$-dimensional lattice L with basis $\mathbf{e}_1, \dots, \mathbf{e}_n, \mathbf{f}$ where $\mathbf{f} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i + \epsilon^{n+1} \mathbf{e}_{n+1}$ and $\{\mathbf{e}_1, \dots, \mathbf{e}_{n+1}\}$ is the standard basis.

The following remarkable result was first stated by Albert Girard (1540–1632) who may have found it on an empirical basis; there is no evidence that he could prove it. The first person to claim to have a proof was Pierre de Fermat (1601–1665). Fermat, however, never published anything mathematical and, while he claimed many discoveries in his correspondence or on the margins of his copy of Diophantus' *Arithmetic* (those marginal notes were later found and published by his son Samuel), there is no trace of proofs, except for one, in his entire extensive surviving correspondence. A century later Leonhard Euler (1707–1783) took great pride in providing proofs of Fermat's theorems, including this one. We give a more recent, devilishly clever proof, based on Minkowski's Theorem and found by Paul Turán (1910–1976).

Exercise* 4.6.7 (Girard-Fermat-Euler)**.** Prove: a prime p can be written as the sum of two squares if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

Hint. Necessity was established in Exercise 4.4.23. For sufficiency, assume $p \equiv 1 \pmod 4$. Then $\left(\dfrac{-1}{p}\right) = 1$ by Exercise 4.5.14 and therefore $(\exists a)(p \mid a^2 + 1)$. Consider the lattice (plane grid) $L \subset \mathbb{Z}^2$ consisting of all integral linear combinations of the vectors $(a, 1)$ and $(p, 0)$. Observe that if $(x, y) \in L$ then $p \mid x^2 + y^2$. Moreover, the area of the fundamental parallelogram of the lattice is p. Apply Minkowski's Theorem to this lattice to obtain a nonzero lattice point (x, y) satisfying $x^2 + y^2 < 2p$.

Chapter 5

Counting

5.1 Binomial coefficients

Exercise 5.1.1. For $n \geq 5$, let $S_n = \binom{5}{5} + \binom{6}{5} + \cdots + \binom{n}{5}$. Prove that

$$S_n = \binom{n+1}{6}.$$

Hint: mathematical induction. Make your proof very simple. You should not need any calculations, just use what we learned in class about binomial coefficients.

Exercise 5.1.2. Prove: if p is a prime number and $1 \leq k \leq p-1$ then p divides the binomial coefficient $\dbinom{p}{k}$.

Exercise 5.1.3. Give closed form expressions (no product symbols or dot-dot-dots) of the binomial coefficients below, using "old" binomial coefficients:

(a) $\dbinom{-1}{k}$

(b) $\dbinom{-1/2}{k}$

where k is a positive integer.

Exercise 5.1.4. Let O_n denote the number of odd subsets of an n-set and E_n the number of even subsets of an n-set. For $n \geq 1$, prove that $O_n = E_n$. Give

(a) a bijective (combinatorial) proof;

(b) an algebraic proof. (Use the Binomial Theorem for the algebraic proof.)

Exercise 5.1.5. Give a closed form expression for

$$\binom{n}{0}+\binom{n}{2}+\binom{n}{4}+\binom{n}{6}+\dots.$$

Exercise$^+$ 5.1.6. Give a closed form expression for

$$\binom{n}{0}+\binom{n}{4}+\binom{n}{8}+\binom{n}{12}+\dots.$$

Hint. Apply the Binomial Theorem to $(1+x)^n$; substitute $1, i, -1, -i$ for x (where $i=\sqrt{-1}$).

Exercise 5.1.7. Prove: $\binom{2n}{n} < 4^n$. Do NOT use Stirling's formula. Your proof should be just one line.

Exercise 5.1.8. Let $n \geq 7$. Count those strings of length n over the alphabet $\{A, B\}$ which contain at least $n-3$ consecutive A's.
Hint. Inclusion–exclusion.

Exercise 5.1.9. Prove: if $1 \leq k \leq n$ then

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

Your proof should be no more than a couple of lines.

Exercise 5.1.10. Prove: if $1 \leq k \leq n$ then

$$\binom{n}{k} < \left(\frac{\mathrm{e}n}{k}\right)^k.$$

Hint. Use the Binomial Theorem and the fact that $(\forall x \neq 0)(\mathrm{e}^x > 1+x)$. (Note that Stirling's formula is of no use; it would only prove things for "large enough n.")

Exercise$^+$ 5.1.11. Prove: if $1 \leq k \leq n$ then

$$\sum_{j=0}^{k}\binom{n}{j} < \left(\frac{\mathrm{e}n}{k}\right)^k.$$

Hint. As in the previous exercise.

Exercise 5.1.12. (a) Evaluate the sum $S_n = \displaystyle\sum_{i=0}^{\infty}\binom{n}{i}2^i$. Your answer should be a very simple closed-form expression (no summation symbols or dot-dot-dots).

(b) Let b_n be the largest term in the sum S_n. Prove: $b_n = \Theta(S_n/\sqrt{n})$.

Exercise 5.1.13. An airline wishes to operate m routes between a given set of n cities. Count the number of possibilities. (A "route" is a pair of cities between which the airline will operate a direct flight. The cities are given, the routes need to be selected. There are no "repeated routes.") Your answer should be a very simple formula.

Exercise 5.1.14. Evaluate the following sums. In each case, your answer should be a simple closed-form expression.

1. $\sum_{i=1}^{n} 4^{n-i}$
2. $\sum_{i=1}^{n} \binom{n}{i} 4^{n-i}$

Exercise 5.1.15. Out of n candidates, an association elects a president, two vice presidents, and a treasurer. Count the number of possible outcomes of the election. (Give a simple expression. State, do not prove.)

Exercise 5.1.16. State your answers as very simple expressions.

1. Count the strings of length 3 (3-letter "words") over an alphabet of n characters.
2. What is the answer to the previous question if no repeated letters are allowed?

Exercise 5.1.17. Evaluate the expression $\binom{0.4}{2}$. Give your answer as a decimal.

Exercise 5.1.18. Pascal's Identity states that $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$. Give a combinatorial proof.

Exercise 5.1.19. We have 5 red beads and 11 blue beads. Count the necklaces that can be made out of these 16 beads. A "necklace" is an arrangement of the beads in a circle. The necklace obtained by rotating the circle does not count as a different necklace. Give a simple expression; do not evaluate.

Exercise 5.1.20. Use the idea of the preceding problem to prove that if a and b are relatively prime then $a+b \mid \binom{a+b}{a}$.

Exercise 5.1.21. Let $a_1, \dots, a_k$ be positive integers. Prove: the least common multiple $L = \text{l.c.m.}\,(a_1, \dots, a_k)$ can be expressed through g.c.d's of subsets of the a_i as follows:

$$L = \prod_{I \subseteq [k]} (\gcd(a_i \,:\, i \in I))^{(-1)^{|I|+1}}.$$

Before attempting to solve this problem for all k, write down the expressions you get for $k = 2$ and $k = 3$ (without the product sign).

5.2 Recurrences, generating functions

Exercise 5.2.1. Let F_n denote the n-th Fibonacci number. ($F_0 = 0, F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$.) Prove: $F_0 + F_1 + \cdots + F_n = F_{n+2} - 1$.

Exercise 5.2.2. Let $a_0 = 3$, $a_1 = 1$, and $a_n = a_{n-1} + a_{n-2}$ ($n \geq 2$) (Fibonacci recurrence with different initial values).

(a) Give a closed-form expression for the generating function $f(x) = \sum_{n=0}^{\infty} a_n x^n$.

(b) Using the generating function, find a closed-form expression for a_n. Show all your work.

Exercise 5.2.3. Let $b_0 = 1$ and $b_n = 3b_{n-1} - 1$ ($n \geq 1$).

(a) (4 points) Give a closed-form expression for the generating function $g(x) = \sum_{n=0}^{\infty} b_n x^n$.

(b) (4 points) Using the generating function, find a closed-form expression for b_n. Show all your work.

Exercise 5.2.4. What is the generating function of each of the following sequences? Give a closed-form expression. Prove your answers.

(a) $a_n = n$.

(b) $b_n = \binom{n}{2}$.

(c) $c_n = n^2$.

(d) $d_n = 1/n!$.

(e) $e_n = 1/n$.

Exercise 5.2.5. If the generating function of the sequence $\{a_n\}$ is $f(x)$, what is the generating function of the sequence $b_n = na_n$? Your answer should be a very simple expression involving $f(x)$ (less than half a line).

Exercise 5.2.6. Let $m_0 = 1$, $m_1 = 2$, and $m_n = m_{n-1} + m_{n-2} + 1$. Express m_n through the Fibonacci numbers. Your expression should be very simple, less than half a line. Do not use generating functions. *Hint.* Tabulate the sequence. Compare with the Fibonacci numbers. Observe the pattern, prove by induction. Watch the subscripts.

Exercise 5.2.7. The sequence $\{a_n\}$ satisfies the recurrence $a_n = 5a_{n-1} - 6a_{n-2}$. Suppose the limit $L = \lim_{n\to\infty} a_n/a_{n-1}$ exists. Determine L.

Exercise 5.2.8. Let the sequence $\{b_n\}$ be defined by the recurrence $b_n = (b_{n-1} + 1)/n$ with initial value $b_0 = 0$. Let $f(x) = \sum_{n=0}^{\infty} b_n x^n$ be the generating function of the sequence. Write a differential equation for f: express $f'(x)$ in terms of $f(x)$ and x. Your expression should be very simple and closed-form.

Exercise 5.2.9. Let r_n be the number of strings of length n over the alphabet $\{A, B\}$ without consecutive A's (so $r_0 = 1$, $r_1 = 2$, $r_2 = 3$). Prove: $r_n \sim c\gamma^n$ where $\gamma = (1+\sqrt{5})/2$ is the golden ratio. Determine the constant c. Prove your answers.

Chapter 6

Graphs and Digraphs

6.1 Graph Theory Terminology

WARNING: There are significant **variations** in **graph theoretic terminology** in the literature. Bear this in mind when using web-sources; please always consult the definitions in this chapter when interpreting a problem or writing a solution. In particular, our terminology significantly **differs from** that of **Rosen's text**. We indicate the main differences below. All concepts below refer to a ("simple") graph $G = (V, E)$.

Exercises. The unmarked exercises are routine, the exercises marked with a "plus" (+) are creative, those marked with an asterisk (*) are challenging; those marked with two asterisks are gems of mathematical ingenuity.

Terminology 6.1.1.

- A **graph** (in Rosen's text: *simple graph*) is a pair $G = (V, E)$ where V is the set of **vertices** and E is the set of **edges**. An **edge** is an unordered pair of **distinct** vertices. We also write $V(G)$ for the set of vertices and $E(G)$ for the set of edges of the graph G. Two vertices joined by an edge are said to be **adjacent**. (So a vertex is never adjacent to itself.) A vertex u and an edge $\{v, w\}$ are **incident** if $u \in \{v, w\}$. Two vertices are **neighbors** if they are adjacent (so they are the two vertices incident with an edge). The **degree** $\deg(v)$ of vertex v is the number of its neighbors. A graph is **regular** of degree r if all vertices have degree r.

- The **complement** $\overline{G}$ of the graph G is the graph $\overline{G} = (V, \overline{E})$ where $\overline{E}$ is the complement of E with respect to the set $\binom{V}{2}$, the set of all unordered pairs of vertices. So $\overline{G}$ has the same set of vertices as G; two distinct vertices are adjacent in $\overline{G}$ if and only if they are not adjacent in G.

- $G = (V, E)$ is a **bipartite graph** is there is a partition of V into two (disjoint) sets, V_1 and V_2 (so $V = V_1 \dot\cup V_2$) such that there are no edges within either set V_i (so every edge joins a vertex in V_1 to a vertex in V_2).

Notation 6.1.2. The **adjacency relation** is denoted $\sim$: we write $u \sim v$ to indicate that u and v are adjacent, and we write $u \sim_G v$ if the graph G needs to be specified.

Exercise 6.1.3. The adjacency relation is symmetric and irreflexive.

Definition 6.1.4 (Isomorphism)**.** An *isomorphism* between the graphs $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \to W$ from V to W which preserves adjacency, i. e., $(\forall x, y \in V)(x \sim_G y \Leftrightarrow f(x) \sim_H f(y)$ The graphs G and H are **isomorphic** if there *exists* a $G \to H$ isomorphism. We denote this circumastance by $G \cong H$.

Exercise 6.1.5. Isomorphism of graphs (understood as the relation of being isomorphic) is an *equivalence relation* on the class of graphs.

Exercise 6.1.6. (a) Draw two non-isomorphic regular graphs with the same number of vertices and the same degree. Make the graphs as small as possible: smallest number of vertices, and given that number of vertices, the smallest number of edges. Prove that your graphs are not isomorphic. Do not prove that they are smallest. (b) Same as part (a) with the additional requirement that both the graphs and their complements be connected.

Notation 6.1.7. Unless expressly stated otherwise, the number of vertices will be denoted by n, and the number of edges by m. If we wish to emphasize the graph G to which the notation refers, we write n_G and m_G for these quantities.

Exercise 6.1.8 (Handshake Theorem)**.** $\sum_{v \in V} \deg(v) = 2m$.

Exercise 6.1.9. Observe: $0 \le m \le \binom{n}{2}$.

Exercise 6.1.10. Observe: $m_G + m_{\overline{G}} = \binom{n}{2}$.

Exercise 6.1.11. A graph is *self-complementary* if it is isomorphic to its complement. (a) Construct a self-complementary graph with 4 vertices. (b) Construct a self-complementary graph with 5 vertices. (c) Prove: if a graph with n vertices is self-complementary then $n \equiv 0$ or 1 (mod 4).

Exercise 6.1.12. (a) Let $b_n = 2^{\binom{n}{2}}$ and $a_n = b_n/n!$. Prove: $\log_2 a_n \sim \log_2 b_n$.

(b) Let $G(n)$ denote the number of non-isomorphic graphs on n vertices.
Prove: $a_n \le G(n) \le b_n$.

(c)* Prove: $G(n) \sim a_n$. *Hint.* Reduce this question to the following: The expected number of automorphisms of a random graph in $1 + o(1)$. (Automorphism = self-isomorphism, i. e., an adjacency-preserving permutation of the set of vertices.)

Terminology 6.1.13 (Paths, cycles, cliques, bipartite cliques, hypercubes)**.**

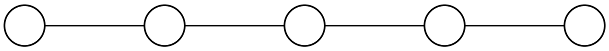


Figure 6.1: P_5, the path of length 4.

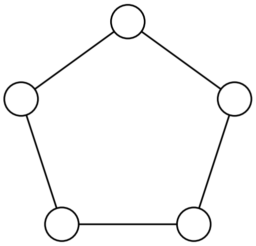


Figure 6.2: C_5, the cycle of length 5.

- The **path** P_n of **length** $n-1$ ($n \geq 1$) edges has n vertices, $V(P_n) = \{v_1, \ldots, v_n\}$, and $n-1$ edges, $E(P_n) = \{\{v_{i-1}, v_i\} \mid 2 \leq i \leq n\}$. We say that this path *connects* v_1 to v_n. See Figure 6.1.

- The **cycle** C_n of **length** n ($n \geq 3$) has n vertices, $V(C_n) = \{v_1, \ldots, v_n\}$, and n edges, $E(C_n) = \{\{v_{i-1}, v_i\} \mid 2 \leq i \leq n\} \cup \{\{v_n, v_1\}\}$. The cycle C_3 is called a **triangle** and C_n is called an n**-cycle.** See Figure 6.2.

- In a **complete graph**, all pairs of vertices are adjacent. A complete graph is also called a **clique**. The complete graph on n vertices (the "n-clique") is denoted by K_n. It has $\binom{n}{2}$ edges. See Figure 6.3.

- The vertices of a **complete bipartite graph** are split into two subsets $V = V_1 \dot\cup V_2$; and $E = \{\{x, y\} : x \in V_1, y \in V_2\}$ (each vertex in V_1 is adjacent to every vertex in V_2, and there are no edges within either V_i). If $k = |V_1|$ and $\ell = |V_2|$ then we obtain the graph $K_{k,\ell}$. This graph has $n = k + \ell$ vertices and $m = k\ell$ edges. A complete bipartite graph is also called a **bipartite clique**. See Figure 6.4.

- The d**-dimensional cube** graph has $n = 2^d$ vertices, labeled by the 2^d $(0,1)$-sequences. Two such sequences, $(a_1, \ldots, a_d)$ and $(b_1, \ldots, b_d)$ ($a_i, b_i \in \{0, 1\}$, are adjacent of they differ in exactly one coordinate, i. e., if $\sum_{i=1}^{d} |a_i - b_i| = 1$.

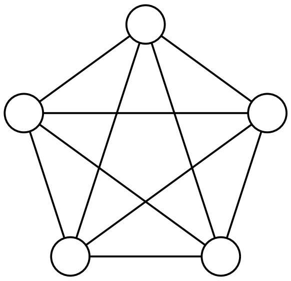


Figure 6.3: The complete graph K_5.

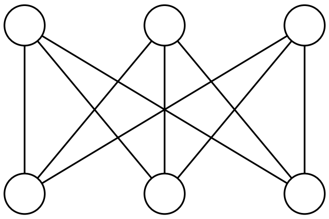


Figure 6.4: The complete bipartite graph $K_{3,3}$.

Exercise 6.1.14. Note that $K_1 \cong P_1 \cong K_{1,0} \cong Q_0$, $K_2 \cong P_2 \cong K_{1,1} \cong Q_1$, $K_3 \cong C_3$, $P_3 \cong K_{2,1}$, $C_4 \cong K_{2,2} \cong Q_2$, $K_{n,0} \cong \overline{K}_n$.

Exercise 6.1.15. Prove: Q_d is bipartite.

Terminology 6.1.16 (Subgraphs)**.**

- The graph $H = (W, F)$ is a **subgraph** of $G = (V, E)$ if $W \subseteq V$ and $F \subseteq E$.
  Notation: $H \subseteq G$.
- $H = (W, F)$ is a **spanning subgraph** of G if $H \subseteq G$ and $V = W$.
- H is an **induced subgraph** of G if $H \subseteq G$ and $(\forall x, y \in W)(x \sim_H y \Leftrightarrow x \sim_G y)$. (So to obtain an induced subgraph, we may delete some vertices and the edges incident with the deleted vertices but no more edges.) Notation: $H = G[W]$. This notation expresses the fact that the graph G and the set $W \subseteq V$ determine H.

Exercise 6.1.17. Observe: (a) Every graph on n vertices is a spanning subgraph of K_n.
(b) Every bipartite graph is a spanning subgraph of a complete bipartite graph.

Exercise 6.1.18. (a) All induced subgraphs of a clique are cliques.
(b) All induced subgraphs of a bipartite clique are bipartite cliques.

Exercise 6.1.19. Let G be a graph with n vertices and m edges. Then G has 2^n induced subgraphs and 2^m spanning subgraphs.

Exercise 6.1.20. Count those spanning subgraphs of K_n which have exactly m edges. Give a simple closed-form expression in terms of binomial coefficients.

Exercise$^+$ 6.1.21. Count the subgraphs of the path P_n. Your answer should be a very simple expression in terms of a familiar sequence.

Exercise 6.1.22. (a) All paths are bipartite. (b) The cycle C_n is bipartite if and only if n is even. (c) Consequently, a bipartite graph cannot contain an odd cycle.

Exercise$^+$ 6.1.23. A graph is bipartite if and only if it contains no odd cycles.

Exercise 6.1.24. Let G be a bipartite graph. Prove: $m \leq \lfloor n^2/4 \rfloor$.

Definition 6.1.25. A graph G is **triangle-free** if it contains no triangles, i. e., $K_3 \not\subseteq G$.

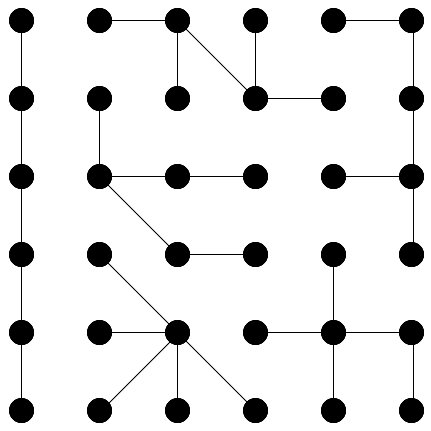


Figure 6.5: The trees on 6 vertices (complete list).

Exercise$^+$ 6.1.26. (Willem Mantel, 1907) Prove: If G is a *triangle-free* graph ($K_3 \not\subseteq G$) then $m \leq \lfloor n^2/4 \rfloor$. Show that this bound is tight for every n. *Hint.* State a lemma about the sum the degrees of a pair of adjacent vertices. Then proceed by induction in increments of 2.

Terminology 6.1.27 (Walks, paths, connected componets, cycles, trees)**.** This is the area where our terminology most differs from Rosen's.

- **walk** (in Rosen: *path*) of length k: a sequence of $k+1$ vertices $v_0, \dots, v_k$ such that v_{i-1} and v_i are adjacent for all i.
- **trail** (in Rosen: *simple path*): a walk without repeated edges.
- **path** in a graph (this all-important concept has no name in Rosen's text): a subgraph that is a path. (Note that the terms "path" and even "simple path" in Rosen's text allow vertices to be repeated.)
- **closed walk** (in Rosen: *circuit* or *cycle*) of length k: a walk $v_0, \dots, v_k$ where $v_k = v_0$.
- **cycle** in a graph (this all-important concept has no name in Rosen's text): a subgraph that is a cycle.
- vertex t is **accessible** from vertex s if there exists an $s \dots t$ walk.
- a graph G is **connected** if every vertex is accessible from each vertex.
- a **tree** is a connected graph without cycles. See Figure 6.5.
- H is a **spanning tree** of G if H is a spanning subgraph of G that is a tree.

Exercise 6.1.28. The accessibility relation is an equivalence relation on V. The equivalence classes are called the **connected components** of the graph.

Exercise 6.1.29. If the vertex t is accessible from the vertex s then there is a path connecting s to t. Consequently, two vertices are in the same connected component if and only if there is a path connecting them.

Exercise 6.1.30. Prove: if a vertex v has odd degree in the graph G then there exists another vertex w, also of odd degree, in the same connected component.

Exercise 6.1.31. Prove that every tree with $n \geq 2$ vertices has at least two vertices of degree 1. *Hint.* Prove that the endpoints of a longest path in a tree have degree 1.

Exercise 6.1.32. Prove that every tree has $n-1$ edges. *Hint.* Induction. Use the preceding exercise.

Exercise 6.1.33. Prove: if G is a connected graph then G has at least $n-1$ edges. Moreover, G is a tree if and only if G is connected and has exactly $n-1$ edges.

Exercise 6.1.34. Draw all non-isomomrphic trees with 7 vertices. Avoid three kinds of mistakes: (a) missing a tree (b) duplicating a tree (drawing two isomorphic trees) (c) drawing a graph that does not belong (it is not a 7-vertex tree). Clearly state the number of trees you found. *Hint.* List the trees in some systematic fashion.

Exercise 6.1.35. Prove: in a tree, all longest paths have a common vertex.

Exercise$^+$ 6.1.36. Let $d_1, \ldots, d_n$ be positive integers such that $\sum_{i=1}^{n} d_i = 2n-2$. Consider those spanning trees of K_n which have degree d_i at vertex i. Count these spanning trees; show that their number is

$$\frac{(n-2)!}{\prod_{i=1}^{n}(d_i-1)!}.$$

Exercise* 6.1.37. (Cayley) The number of spanning trees of K_n is n^{n-2}. *Hint.* This amazingly simple formula is in fact a simple consequence of the preceding exercise. Use the Multinomial Theorem.

Exercise 6.1.38. Let $t(n)$ denote the number of non-isomorphic trees on n vertices. Use Cayley's formula to prove that $t(n) > 2.7^n$ for sufficiently large n (i. e., $(\exists n_0)(\forall n > n_0)(t(n) > 2.7^n)$).

Exercise$^+$ 6.1.39. Find a constant C such that $t(n) \leq C^n$.

Exercise 6.1.40. Count the 4-cycles in the complete bipartite graph $K_{m,n}$. (You need to count those subgraphs which are isomorphic to C_4.) (*Comment.* Check two small cases: $K_{2,2} \cong C_4$ has exactly one 4-cycle, and $K_{2,3}$ has three 4-cycles. Your answer should be a very simple formula, consistent with these initial values.)

Exercise$^+$ 6.1.41. (Kővári–Sós–Turán) Prove: if G has no 4-cycles ($C_4 \not\subseteq G$) then $m = O(n^{3/2})$. Show that this bound is tight (apart from the constant implied by the big-Oh notation). *Hint.* Count the paths of length 2 in G in two ways.

Terminology 6.1.42 (Cliques, distance, diameter, chromatic number)**.**

- A **k-clique** in G is a subset $W \subseteq V$ of size k such that the induced subgraph $G[W]$ is a clique (i. e., the vertices in W are pairwise adjacent). The **clique number** of G, denoted $\omega(G)$, is the size of the largest clique in G.

- An **independent set** or **anti-clique** of size k in G is a set of k pairwise non-adjacent vertices. In other words, an independent set in G is a clique in $\overline{G}$. The **independence number** of G, denoted $\alpha(G)$, is the size of the largest independent set in G. In other words, $\alpha(G) = \omega(\overline{G})$.

- The **distance** $\text{dist}(x, y)$ between two vertices $x, y \in V$ is the length of a shortest path between them. If there is no path between x and y then their distance is said to be infinite: $\text{dist}(x, y) = \infty$.

- The **diameter** of a graph is the maximum distance between all pairs of vertices. So if a graph has diameter d then $(\forall x, y \in V)(\text{dist}(x, y) \le d)$ and $(\exists x, y \in V)(\text{dist}(x, y) = d)$.

- The **girth** of a graph is the length of its shortest cycle. If a graph has no cycles then its girth is said to be infinite.

- A **legal k-coloring** of a graph is a function $c : V \to [k] = \{1, \ldots, k\}$ such that adjacent vertices receive different colors, i. e., $u \sim v \Rightarrow c(u) \neq c(v)$. A graph is **$k$-colorable** if there exists a legal k-coloring. The **chromatic number** of G, denoted $\chi(G)$, is the smallest k such that G is k-colorable.

- A **Hamilton cycle** is a cycle of length n, i. e., a cycle that passes through all vertices. G is **Hamiltonian** if it has a Hamilton cycle.

- A **Hamilton path** is a path of length $n - 1$, i. e., a path that passes through all vertices.

Exercise 6.1.43 (Girth)**.** Verify: for $n = 3$, the clique K_n has girth 3; the icosahedron graph has girth 3; the bipartite clique $K_{k,\ell}$ has girth 4 if $k, \ell \ge 2$; the $k \times \ell$ grid (Figure 6.6) has girth 4, assuming $k, \ell \ge 2$; the d-dimensional cube Q_d has girth 4 ($d \ge 2$); the Petersen graph (Figure 6.8) and the dodecahedron graph have girth 5; the incidence graphs of finite projective planes have girth 6; trees have infinite girth.

Exercise* 6.1.44. (a) For every $g \ge 3$ find a trivalent graph (a regular graph of degree 3) of girth $\ge g$. (b) For every $g \ge 3$ and $d \ge 3$ find a d-regular graph of girth $\ge g$.

Exercise 6.1.45. Verify: the diameter (a) of P_n is $n - 1$; (b) of C_n is $\lfloor n/2 \rfloor$; (c) of K_n is 1; (d) of $K_{k,\ell}$ is 2 (assuming $k\ell \ge 2$); (e) of Q_d is d.

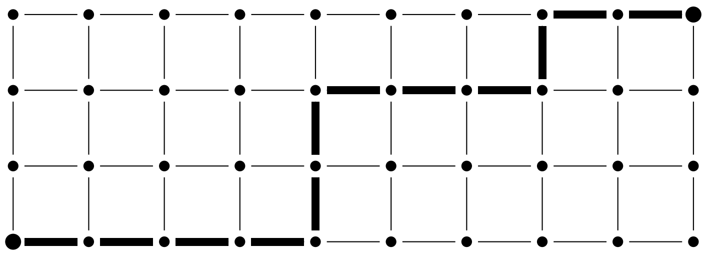


Figure 6.6: The 4×10 grid, with a shortest path between opposite corners highlighted.

Exercise 6.1.46. (a) Disprove the following statement: "the diameter of a graph is the length of its longest path." (b) For every n, find the maximum ratio between the length of the longest path and the diameter. (c) Prove that the statement is true for trees: the diameter is the length of the longest path.

Exercise 6.1.47. The $k \times \ell$ grid has $k\ell$ vertices (Figure 6.6). Count its edges.

Exercise 6.1.48. Verify that the diameter of the $k \times \ell$ grid is $k + \ell - 2$.

Exercise 6.1.49. Let u and v be two opposite corners of the $k \times \ell$ grid. Count the shortest paths between u and v. Your answer should be a very simple expression in terms of binomial coefficients.

Exercise 6.1.50. A graph is bipartite if and only if it is 2-colorable.

Exercise 6.1.51. We color the vertices of a bipartite graph G red and blue (legal coloring). Assume G has 30 red vertices (all other vertices are blue). Suppose each red vertex has degree 6 and each blue vertex has degree 5. What is the number of blue vertices? Prove your answer.

Exercise 6.1.52. Let us pick 3 distinct vertices at random in a bipartite graph G with n vertices. Prove that the probability that we picked an independent set is $\geq 1/4 - o(1)$ (as $n \to \infty$).

Exercise 6.1.53. For every $n \geq 1$, name a graph with n vertices, at least $(n^2 - 1)/4$ edges, and no cycles of length 5.

Exercise 6.1.54. Prove: if every vertex of a graph has degree $\leq d$ then the graph is $d+1$-colorable (i. e., $\chi(G) \leq d + 1$).

Exercise 6.1.55. For every n, construct a 2-colorable graph with n vertices such that every vertex has degree $\geq (n-1)/2$. (Moral: low degree is a sufficient but not a necessary condition of low chromatic number.)

Exercise 6.1.56. (Chromatic number vs. independence number) Prove: $\alpha(G)\chi(G) \geq n$.

Exercise 6.1.57. $\chi(G) \geq \omega(G)$. So graphs that contain large cliques have large chromatic number.

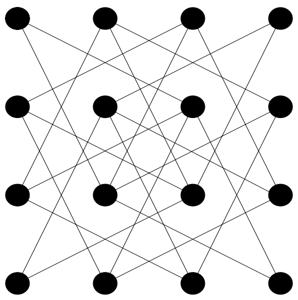


Figure 6.7: Graph of knight moves on a 4 × 4 chessboard

But are large cliques the only reason for large chromatic number?

Exercise 6.1.58. (a) Find the smallest graph G such that $\chi(G) \neq \omega(G)$. (b) Find the smallest graph G such that $\chi(G) \geq 4$ and $\omega(G) = 3$.

Exercise$^+$ 6.1.59. Construct a graph G on 11 vertices such that G is triangle-free ($K_3 \not\subseteq G$) and G is NOT 3-colorable. Prove that your graph has the stated properties. *Hint.* Draw your graph so that it has a rotational symmetry of order 5 (rotation by $2\pi/5$ does not change the picture).

Exercise* 6.1.60. Prove: $(\forall k)(\exists G)(\chi(G) \geq k$ and G is triangle-free.)

The following celebrated result is one of the early triumphs of the "Probabilistic Method." You can find the elegant proof in the book by Alon and Spencer.

Theorem 6.1.61. (Erdős, 1959) *Prove:* $(\forall k, g)(\exists G)(\chi(G) \geq k$ *and* G *has girth* $\geq g.)$

Exercise 6.1.62. Count the Hamilton cycles in the complete graph K_n.

Exercise 6.1.63. Count the Hamilton cycles in the complete bipartite graph $K_{r,s}$. (Make sure you count each cycle only once – note that $K_{2,2}$ has exactly one Hamilton cycle.)

Exercise 6.1.64. Prove that all grid graphs have a Hamilton path.

Exercise 6.1.65. Prove: the $k \times \ell$ grid is Hamiltonian if and only if $k, \ell \geq 2$ and $k\ell$ is even. (Your proofs should be very short, only one line for non-Hamiltonicity if $k\ell$ is odd.)

Exercise 6.1.66. Prove that the dodecahedron graph is Hamiltonian. (Lord Hamilton entertained his guests with this puzzle; hence the name.)

Exercise 6.1.67. (a) Prove: the graph of the knight's moves on a 4 × 4 chessboard (Figure 6.7) has no Hamilton path. Find an "Ah-ha!" proof: just "one line" after the following Lemma.

(b) Lemma. If a graph has a Hamilton path then after deleting k vertices, the remaining graph has $\leq k + 1$ connected components.

Exercise 6.1.68. We have a standard (8×8) chessboard and a set of 32 dominoes such that each domino can cover two neighboring cells of the chessboard. So the chessboard can be covered with the dominoes. Prove: if we remove the top left and the bottom right corner cells of the chessboard, the remaining 62 cells cannot be covered by 31 dominoes. Find an "Ah-ha!" proof (elegant, no case distinctions.)

Exercise 6.1.69. A mouse finds a $3 \times 3 \times 3$ chunk of cheese, cut into 27 blocks (cubes), and wishes to eat one block per day, always moving from a block to an adjacent block (a block that touches the previous block along a face). Moreover, the mouse wants to leave the center cube last. Prove that this is impossible. Find two "Ah-ha!" proofs; one along the lines of the solution of Exercise 6.1.67, the other inspired by the solution of Exercise 6.1.68.

Exercise 6.1.70. Prove that the Petersen graph (Figure 6.8) is not Hamiltonian; its longest cycle has 9 vertices. (No "Ah-ha!" proof of this statement is known.)

Exercise 6.1.71. Prove: if G is regular of degree r and G has girth ≥ 5 then $n \geq r^2 + 1$. (n is the number of vertices.) Show that $n = r^2 + 1$ is possible for $r = 1, 2, 3$.

Exercise 6.1.72. (a) Prove: if a graph G with n vertices is regular of degree r and has diameter 2 then $n \leq r^2 + 1$.

(b) Prove that if G is as in part (a) and $n = r^2 + 1$ then G has girth 5.

(c) Show that there exists a graph G satisfying the conditions of part (a) and the equation $n = r^2 + 1$ if $r = 2$ or $r = 3$ (what is the name of your graph?). *Remark.* $n = r^2 + 1$ is possible also if $r = 7$ (the "Hoffman–Singleton graph"). It is known **(Hoffman–Singleton, 1960)** that the only values of r for which $n = r^2 + 1$ is conceivable are $2, 3, 7$, and 57. The proof is one of the gems of the applications of linear algebra (the Spectral Theorem) to graph theory. The question whether $r = 57$ can actually occur remains open.

Exercise 6.1.73. An *automorphism* of the graph G is a $G \to G$ isomorphism. (a) Count the automorphisms of K_n, C_n, P_n, Q_n. (b)$^+$ Show that the dodecahedron has 120 automorphisms. (c)$^+$ Show that the Petersen graph has 120 automorphisms.

6.2 Planarity

A *plane graph* is a graph drawn in the plane so that the lines (curves) representing the edges do not intersect (except at their end vertices). A graph is *planar* if it admits a plane drawing; such plane drawings are the *plane representations* of the graph. Of course a planar graph may also have drawings that are not plane graphs (e. g., K_4 is a planar graph - a plane representation is a regular triangle with its center, with their conecting straight line segments; a drawing of K_4 which is not a plane graph is the square with all sides and diagonals–see Figure 6.10).

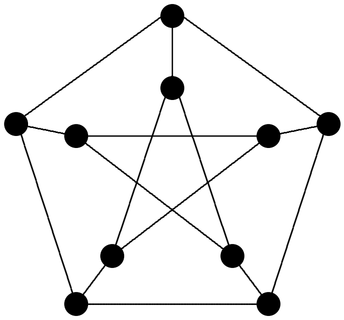


Figure 6.8: The Petersen graph.

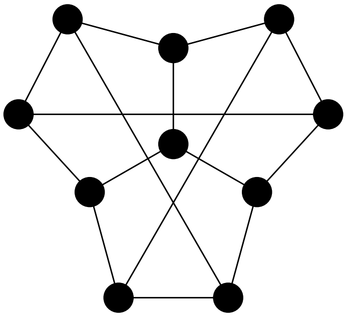


Figure 6.9: Is this graph isomorphic to Petersen's?

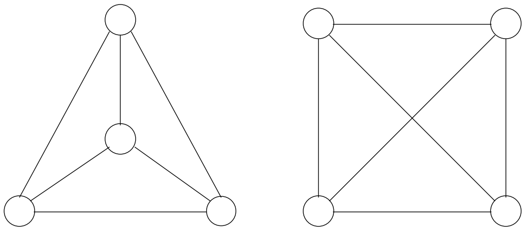


Figure 6.10: K_4 drawn two different ways. Only one is a plane graph.

The *regions* of a plane graph are the regions into which the drawing divides the plane; so two points of the plane belong to the same region if they can be connected so that the connecting line does not intersect the drawing. Note that the infinite "outer region" counts as a region.

WARNING: It is incorrect to speak of regions of a *planar* graph; only a *plane* graph has regions. A planar graph may have many inequivalent plane representations; the sizes of the regions may depend on the representation.

Exercise 6.2.1. Prove: every plane representation of a tree has just one region. *Hint.* Induction (use the fact that the tree has a vertex of degree 1).

We need the following, highly nontrivial result.

Theorem 6.2.2. (Jordan's Curve Theorem) *Every plane representation of a cycle has two regions.*

Exercise 6.2.3. (Euler's formula) For a connected plane graph, let n, m, r denote the set of vertices, edges, and regions, respectively. Then $n - m + r = 2$. *Note* that this statement includes Jordan's Curve Theorem and the exercise before that. *Hint.* Induction on m. Unless the graph is a tree, delete an edge contained in a cycle; verify that this reduces the number of regions by 1. Trees are the base case.

Exercise 6.2.4. Verify that the Platonic solids satisfy Euler's formula.

Exercise 6.2.5. Let r_k denote the number of k-sided regions of a plane graph. (In a plane graph, an edge has two sides, and it is possible that both sides are incident with the same region. In such a case this edge contributes 2 to the number of sides of the region. See Figure 6.11.) Prove: $\sum_{k=3}^{n} r_k = 2m$.

Exercise 6.2.6. Prove: in a plane graph, $3r \leq 2m$.

Exercise 6.2.7. Prove: in a plane graph without triangles, $2r \leq m$.

Exercise 6.2.8. Prove: a planar graph with $n \geq 3$ vertices has $m \leq 3n - 6$ edges. *Hint.* Use Euler's formula and the inequality $3r \leq 2m$.

Exercise 6.2.9. Prove: a triangle-free planar graph with $n \geq 3$ vertices has $m \leq 2n - 4$ edges. *Hint.* Use Euler's formula and the inequality $2r \leq m$.

Exercise 6.2.10. Prove: the graphs K_5 and $K_{3,3}$ are not planar. *Hint.* Use the preceding two exercises.

Definition 6.2.11. A **subdivision** of a graph is obtained by subdividing some of its edges by new vertices. For instance, the cycle C_n is a subdivision of the triangle C_3; the path P_n is a subdivision of an edge. Two graphs are **homeomorphic** if both of them is a subdivision of the same graph. For instance, all cycles (including C_3) are homeomorphic. Homeomorphic planar graphs have identical plane drawings.

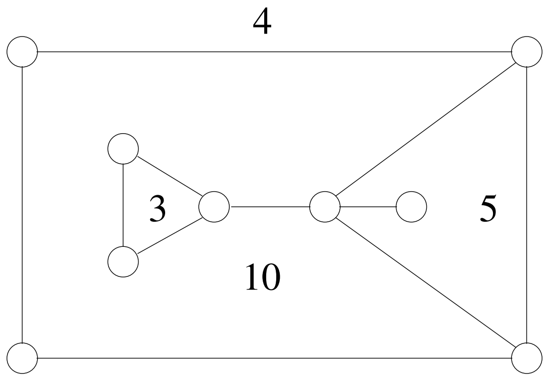


Figure 6.11: The numbers indicate the number of sides of each region of this plane graph.

Kuratowski's celebrated theorem gives a *good characterization* of planarity.

Theorem 6.2.12. *A graph is planar if and only if it does not contain a subgraph homeomorphic to $K_{3,3}$ or K_5.*

The two minimal non-planar graphs, $K_{3,3}$ and K_5, are referred to as the **Kuratowski graphs.**

Exercise 6.2.13. Draw a BIPARTITE graph G which is NOT planar and does NOT contain a subdivision of $K_{3,3}$. Make a clean drawing; your graph should have no more than 20 edges. Prove that your graph has all the required properties.

Exercise 6.2.14. Prove: (a) If a connected graph G has n vertices and $n+2$ edges then G is planar. (b) Show that for every $n \geq 6$, statement (a) becomes false if we replace $n+2$ by $n+3$. (You must construct an infinite family of counterexamples, one graph for each $n \geq 6$.)

Exercise 6.2.15. Prove that every planar graph has a vertex of degree ≤ 5. *Hint.* $m \leq 3n-6$.

Exercise 6.2.16. Prove that every planar graph is 6-colorable. *Hint.* Induction, using the preceding exercise.

The famous **4-Color Theorem** of Appel and Haken asserts that every planar graph is 4-colorable. The proof considers hundreds of cases; no "elegant" proof is known.

Exercise 6.2.17. Prove: if a planar graph G has n vertices then $\alpha(G) \geq n/6$. (Recall that $\alpha(G)$ denotes the maximum number of independent vertices in G.) *Hint.* Use the preceding exercise.

Prove that every triangle-free planar graph has a vertex of degree ≤ 3. *Hint.* $m \leq 2n-4$.

Exercise 6.2.18. Prove that every triangle-free planar graph is 4-colorable.

6.3 Ramsey Theory

The Erdős–Rado **arrow symbol** $n \to (k, \ell)$ denotes the statement that every graph on n vertices either has a clique of size $\geq k$ or an independent set of size $\geq \ell$. In other words, if we color the edges of K_n red and blue, there will either be an all-red K_k or an all-blue K_ℓ.

Exercise 6.3.1. Prove: (a) $6 \to (3, 3)$; (b) $5 \not\to (3, 3)$ (c) $n \to (n, 2)$.

Exercise 6.3.2. (Erdős–Szekeres, 1933)

$$\binom{r+s}{r} \to (r+1, s+1).$$

Hint. Induction on $r + s$.

Exercise 6.3.3. The Erdős–Szekeres theorem tells us that $10 \to (4, 3)$. Strengthen this: prove $9 \to (4, 3)$.

Exercise 6.3.4. Prove: $n \to (k, k)$ where $k = \lceil \log_2 n/2 \rceil$.

Exercise 6.3.5. Define and prove: $17 \to (3, 3, 3)$.

6.4 Digraph Terminology

Definition 6.4.1. A **directed graph** (*digraph*, for short), is a pair $G = (V, E)$, where V is the set of "vertices" and E is a set of ordered pairs of vertices called "edges:" $E \subseteq V \times V$.

Exercise 6.4.2. If G has n vertices and m edges then $m \leq n^2$.

Definition 6.4.3 (Adjacency). We say that u is *adjacent* to v, denoted $u \to v$, if $(u, v) \in E$. *Self-adjacency* may occur; an edge $(u, u) \in E$ is called a **loop.**

Definition 6.4.4 (Representation of graphs as digraphs)**.** "Graphs," also referred to as **undirected graphs**, can be represented as digraphs by introducing a pair of directed edges, (u, v) and (v, u), for every undirected edge $\{u, v\}$ of a graph. (So the digraph G corresponding to the graph G_0 has twice as many edges as G_0.) Note that the digraph representing a graph will have no **loops** (self-adjacencies) ($v \not\to v$) (in other words, the adjacency relation is *irreflexive.*

Definition 6.4.5. The **converse** of a digraph $G = (V, E)$ is the digraph $G^{tr} = (V, E^{tr})$ where E^{tr} consists of all edges of G reversed: $E^{tr} = \{(v, u) : (u, v) \in E\}$. Note that G is **undirected** if and only if $G = G^{tr}$. – The superscript "tr" refers to "transpose," for a reason to be clarified below.

Definition 6.4.6 (Orientations of a graph). Let $G_0 = (V, E_0)$ be a graph. We say that the digraph $G = (V, E)$ is an **orientation** of G_0 if (i) $E_0 = \{\{u, v\} \mid (u, v) \in E\}$, and for each edge $\{u, v\} \in E_0$, exactly one of (u, v) and (v, u) belongs to E. In other words, we put an arrow on every edge of $G_)$.

Exercise 6.4.7. If a graph has m edges then it has 2^m orientations.

Definition 6.4.8. Tournaments are orientations of complete graphs.

So in a tournament $G = (V, E)$, for every pair of vertices $u, v \in V$, exactly one of the following holds: (a) $u = v$; (b) $u \to v$; (c) $v \to u$. We can think of the vertices of a tournament as players in a round-robin tournament without ties or rematches. Each player plays against every other player exactly once; $u \to v$ indicates that player u beat player v.

Exercise 6.4.9. Count the tournaments on a given set of n vertices. Is the similarity with the number of graphs a coincidence?

Definition 6.4.10 (Neighbors). If $u \to v$ in a digraph then we say that v is an **out-neighbor** or **successor** of u; and u is an **in-neighbor** or **predecessor** of v.

Definition 6.4.11. Degrees. The **out-degree** $\deg^+(v)$ of vertex v is the number of its out-neighbors; the **in-degree** $\deg^-(v)$ of v is the number of its in-neighbors.

Exercise 6.4.12 (Directed Handshake Theorem)**.** Prove: if the digraph $G = (V, E)$ has n vertices and m edges then

$$\sum_{v \in V} \deg^+(v) = \sum_{v \in V} \deg^-(v) = m.$$

Exercise 6.4.13. Prove: if every vertex of a digraph G has the same out-degree d^+ and the same in-degree d^- then $d^+ = d^-$.

Definition 6.4.14. An **isomorphism** between the digraphs $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \to W$ from V to W which preserves adjacency, i. e., $(\forall x, y \in V)(x \to_G y \Leftrightarrow f(x) \to_H f(y)$. Two digraphs are **isomorphic** if there *exists* an isomorphism between them. This circumstance is denoted $G \cong H$.

Terminology 6.4.15 (Directed walks, paths, cycles)**.**

- **(directed) walk** (in Rosen's text: *path*) of length k: a sequence of $k + 1$ vertices, $v_0, \ldots, v_k$, such that $(\forall i)(v_{i-1} \to v_i)$.
- **(directed) trail** (in Rosen: *simple path*): a walk without repeated edges.
- **(directed) path:** (this all-important concept has no name in Rosen): a walk without repeated vertices. (Note that the terms "path" and even "simple path" in Rosen allow vertices to be repeated.) $\vec{P}_{k+1}$ denotes a directed path of length k (it has $k + 1$ vertices)
- **closed (directed) walk** (in Rosen: *circuit* or *cycle*) of length k: a (directed) walk $v_0, \ldots, v_k$ where $v_k = v_0$.
- **(directed) cycle of length** k or k**-cycle**: (this all-important concept has no name in Rosen): a closed walk of length k with no repeated vertices except that $v_0 = v_k$. Notation: $\vec{C}_k$.
- a vertex v is **accessible** from a vertex u if there exists a $u \to \cdots \to v$ directed path.

Exercise 6.4.16. (a) Prove: If there is a directed walk from vertex s to vertex t then there is a directed path from s to t. (b) Prove that accessibility is a transitive relation.

Exercise 6.4.17. Prove that the relation "u and v are mutually accessible from each other" is an **equivalence relation** on the set of vertices of the digraph G.

Terminology 6.4.18.

- The **strong components** of G are the equivalence classes of this relation, i. e., the *maximal* subsets of the vertex set consisting of mutually accessible vertices. The vertex set of G is the disjoint union of the strong components. In other words, **each vertex belongs to exactly one strong component.** The vertices u and v **belong to the same strong component** if and only if they are mutually accessible from each other.
- a digraph G is **strongly connected** if there is a (directed) path between each pair of vertices, i. e., all vertices belong to the same strong component. (There is just one strong component.)
- an *undirected walk (path, cycle, etc.)* in a digraph is a walk (path, cycle, etc.) in the undirected graph obtained by ignoring orientation.

- a digraph is **weakly connected** if there is an undirected path between each pair of vertices.

Exercise 6.4.19. Prove that a weakly connected digraph has $\geq n-1$ edges; a strongly connected digraph has $\geq n$ edges.

Exercise$^+$ 6.4.20. Prove: if $(\forall v \in V)(\deg^+(v) = \deg^-(v))$ and G is weakly connected then G is strongly connected.

Terminology 6.4.21 (Hamiltonicity)**.**

- A **Hamilton cycle** in a digraph is a (directed) cycle of length n, i. e., a cycle that passes through all vertices. G is **Hamiltonian** if it has a Hamilton cycle.
- A **Hamilton path** in a digraph is a (directed) path of length $n-1$, i. e., a path that passes through all vertices.

Exercise 6.4.22. Prove that every tournament has a Hamilton path.

Exercise$^+$ 6.4.23. Prove that every strongly connected tournament is Hamiltonian.

Terminology 6.4.24 (DAGs, topological sort)**.**

- A DAG **(directed acyclic graph)** is a digraph with no directed cycles.
- A **topological sort** of a digraph is an ordering of its vertices such that all edges go "forward:" if $u \to v$ then u precedes v in the ordering.

Exercise 6.4.25. Prove that a digraph G has a topologically sort if and only if G is a DAG. – Note that this is a **good characterization**: the existence of an object (topological sort) is shown to be equivalent to the nonexistence of another (directed cycle).

Exercise 6.4.26. If $V = \{1, 2, \dots, n\}$ and $u \to v$ means $u \neq v$ and $u \mid v$ (u divides v) then the natural ordering of integers is a topological sort; but it is not the only possible topological sort of this digraph.)

Exercise 6.4.27. Prove that the "divisibility digraph" described in the preceding exercise has at least $\lfloor n/2 \rfloor!$ topological sorts.

Exercise 6.4.28. Prove that for every n, there exists exactly one tournament (up to isomorphism) which is a DAG.

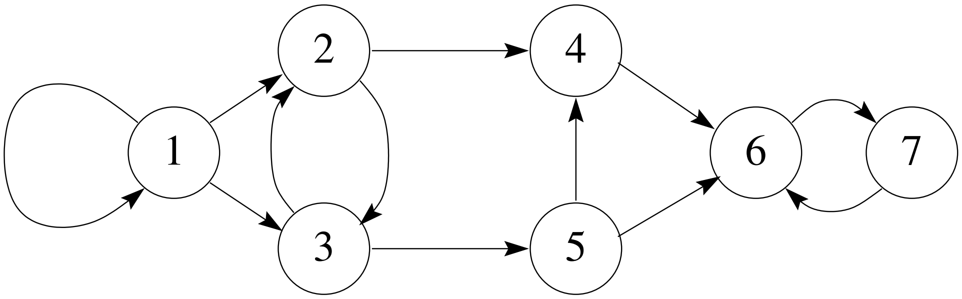
Definition 6.4.29 (Adjacency matrix)**.** Let $G = (V, E)$ be a digraph; assume $V = [n] = \{1, 2, \dots, n\}$. Consider the $n \times n$ matrix $A_G = (a_{ij})$ defined as follows: $a_{ij} = 1$ if $i \to j$; and $a_{ij} = 0$ otherwise. A_G is the **adjacency matrix** of G.

Exercise 6.4.30. Prove: The adjacency matrix of the G^{tr} (the converse of G) is A_G^{tr} (the transpose of A_G. In particular, the digraph G is undirected if and only if A_G is a symmetric matrix.

Exercise$^+$ 6.4.31. (Counting walks) For $k \geq 0$, let a_{ijk} denote the **number of directed walks** of length k from vertex i to vertex j. Consider the matrix $A_G(k)$ which has a_{ijk} as its entry in row i, column j. Prove: $A_G(k) = A_G^k$. *Hint.* Induction on k.

Exercise 6.4.32. Let T be a tournament with n vertices. Prove: if all vertices have the same out-degree then n is odd.

Exercise 6.4.33. List the strong components of the digraph in the figure below. State the number of strong components. Recall that two vertices x and y belong to the same strong component if either $x = y$ or there exists $x \to y$ and $y \to x$ directed walks. The strong components are the equivalence classes of this equivalence relation, so each strong component is either a single vertex or a maximal strongly connected subgraph.



Exercise 6.4.34. Let $p_1, \dots, p_k$ be distinct prime numbers and let $n = \prod_{i=1}^{k} p_i$. Let D denote the set of positive divisors of n.

1. Determine $|D|$ (the size of D). (Your answer should be a very simple formula.)
2. We define a digraph G with vertex set $V(G) := D$ by setting $i \to j$ if $j \mid i$ and i/j is a prime number ($i, j \in D$). Determine the number of directed paths from n to 1 in G. (Again, your answer should be a very simple formula.)
3. Prove that this digraph is self-converse (isomorphic to the digraph obtained by reversing all arrows). (You need to state a bijection $f : D \mapsto D$ which reverses all arrows. You should define f by a very simple formula.)

Definition 6.4.35. Let v be a vertex in a directed graph. The *period* of v is defined as the gcd of the lengths of all closed walks containing v.

Exercise 6.4.36. Let G be a directed graph. Prove: if $v, w \in V$ are two vertices in the same strong component of G then their periods are equal.

6.4.1 Paradoxical tournaments, quadratic residues

Let p be a prime. An integer z is a **quadratic residue** modulo p if $z \not\equiv 0 \pmod p$ and $(\exists x)(x^2 \equiv z \pmod p)$.

Exercise 6.4.37. List the quadratic residues modulo 5 and modulo 7.

Exercise 6.4.38. Prove that if p is an odd prime then the number of non-congruent quadratic residues modulo p is $(p-1)/2$.

Exercise$^+$ 6.4.39. Prove: -1 is a quadratic residue mod p if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

Paley graphs/tournaments. Let p be an odd prime. Let $V = \{0, 1, \dots, p-1\}$. Let us set $u \to v$ if $u - v$ is a quadratic residue mod p. $(0 \le u, v \le p-1.)$

Exercise 6.4.40. Prove: the preceding construction defines a tournament (the **Paley tournament**) if $p \equiv -1 \pmod 4$; and it defines a graph (the **Paley graph**) if $p \equiv 1 \pmod 4$.

A digraph is **self-converse** if it is isomorphic to its converse.

Exercise$^+$ 6.4.41. Prove: (a) The Paley tournaments are self-converse. (b) The Paley tournaments are self-complementary.

Exercise* 6.4.42. (Erdős.) We say that a tournament is k**-paradoxical** if to every k players there exists a player who beat all of them. Prove that if $n > 2k^2 2^k$ then there exists a k-paradoxical tournament on n vertices. *Hint.* Use the probabilistic method: prove that *almost all tournaments* are k-paradoxical.

Exercise** 6.4.43. (Graham – Spencer) If p is a prime, $p \equiv -1 \pmod 4$ and $p > 2k^2 4^k$ then the Paley tournament on p vertices is k-paradoxical. *Hint.* The proof uses André Weil's character sum estimates.

Chapter 7

Finite Probability Spaces

Part of chapter from
László Babai: "Discrete Mathematics" (Lecture notes, 2003, 2020, 2021)
Last updated December 30, 2022.

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 Finite probability space, events

Definition 7.2.1 (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

$$(\forall a \in \Omega)\left(f(a) = \frac{1}{|\Omega|}\right)\,. \tag{7.5}$$

Definition 7.2.2. 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 7.2.3. 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 $0 \leq p \leq 1$. Let us define the function $\Pr_p : \mathbb{B}^n \to \mathbb{R}$ by the formula

$$\Pr_p(s) = p^{k(s)} \cdot (1-p)^{n-k(s)}\,. \tag{7.6}$$

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 7.2.4. (a) Prove that Eq. (7.6) 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 7.2.5 (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

$$\Pr(A) := \sum_{a \in A} \Pr(a)\,. \tag{7.7}$$

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

Exercise 7.2.6. Prove: $\Pr(\emptyset) = 0$ and $\Pr(\Omega) = 1$.

Definition 7.2.7. The **trivial events** are those with probability 0 or 1.

Exercise 7.2.8. Prove: the number of trivial events is a power of 2.

Exercise 7.2.9. In a uniform probability space, calculation of probabilities amounts to counting:

$$\Pr(A) = \frac{|A|}{|\Omega|}\,. \tag{7.8}$$

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 7.2.10 (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 7.2.11 (Coin flips)**.** Let $0 \le p \le 1$. Consider the probability space $(\mathbb{B}^n, \Pr_p)$ where the probability distribution $\Pr_p$ is defined by Eq. (7.6). (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$^+$ 7.2.12. (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 7.2.13 (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 7.2.14 (Modular equation)**.** Let $A, B \subseteq \Omega$ be events. Prove:

$$\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B). \tag{7.9}$$

Definition 7.2.15. 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 7.2.16 (Union bound)**.** Let $A_1, \ldots, A_k \subseteq \Omega$ be events. Prove:

$$\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^{k} \Pr(A_i). \tag{7.10}$$

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

7.3 Conditional probability, probability of causes

Definition 7.3.1 (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

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}\,. \tag{7.11}$$

(B is the *condition.*)

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

Exercise 7.3.2. 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

$$\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)\,. \tag{7.12}$$

Exercise 7.3.3 (Bayes's equation)**.**

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\cdot\Pr(B)}{\Pr(A)}\,. \tag{7.13}$$

Exercise 7.3.4. Prove: $\Pr(A \cap B \cap C) = \Pr(A \mid B \cap C)\Pr(B \mid C)\Pr(C)$.

Exercise 7.3.5. 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 7.3.6. A **partition** of Ω is a set of pairwise disjoint events $H_1,\dots,H_k$ of positive probability, covering Ω:

$$\Omega = H_1 \cup \ldots \cup H_k, \qquad H_i \cap H_j = \emptyset, \qquad (\forall i)(\Pr(H_i) \neq 0)\,. \tag{7.14}$$

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

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A \mid H_i)\Pr(H_i). \tag{7.15}$$

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

Exercise 7.3.8 (Probability of causes)**.** 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 (i. e., the numerator and the denominator are relatively prime).

(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).

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

Definition 7.4.1. 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 7.4.2. Assume $\Pr(B) > 0$. Then A and B are independent $\iff \Pr(A \mid B) = \Pr(A)$.

Definition 7.4.3. The **complement** of the event A is the event $\overline{A} = \Omega \setminus A$.

Exercise 7.4.4 (Independence of complement)**.** If A and B are independent events then $\overline{A}$ and B are also independent.

Exercise 7.4.5 (Independence of a trivial event)**.** Let A be any event and B a trivial event. Show that A and B are independent.

Exercise 7.4.6. 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 7.4.7 (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 7.4.8. 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 7.4.9. Assume there exist two nontrivial independent events in our probability space. Prove: $|\Omega| \geq 4$.

Definition 7.4.10 (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 7.4.11. 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 7.4.12. Pick a number x at random from $[n] = \{1, \dots, n\}$. 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 7.4.13. Let $A \subseteq B \subseteq \Omega$. If A and B are independent then one of them is trivial.

Exercise 7.4.14. 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 7.4.15. 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.

7.5 Independence of multiple events

Definition 7.5.1. Events $A_1, \dots, A_k$ are **pairwise independent** if $(\forall i \neq j)(A_i$ and A_j are independent).

Definition 7.5.2 (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 7.5.3 (Independence of complement)**.** If A, B, and C are independent events then $\overline{A}, B$ and C are also independent.

Exercise 7.5.4 (Independence of trivial event)**.** If A, B are independent and C is a trivial event then A, B, C are indepenedent.

Exercise 7.5.5 (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 7.5.6. Assume there exist three nontrivial independent events in our probability space. Prove: $|\Omega| \geq 8$.

Exercise 7.5.7. (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 7.5.8. A **balanced event** is an event of probability 1/2.

Exercise 7.5.9. (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 7.5.2 suggests the following inductive definition. A k**-subset** is a subset of size k.

Definition 7.5.10. We say that the events $A_1, \dots, A_n$ 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 7.5.11 (Independence: inductive definition)**.** Let $k \geq 3$. We say that the events $A_1, \dots, A_k$ are independent if

(i) $A_1, \dots, A_k$ 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. 7.5.11 will take effect for $k \geq 2$.

Next we give an alternative, non-inductive definition.

Definition 7.5.12 (Independence: explicit definition)**.** Events $A_1, \dots, A_k$ are **independent** if for all subsets $I \subseteq [k]$ we have

$$\Pr\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \Pr(A_i)\,. \tag{7.16}$$

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 7.5.13 (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

$$\bigcap_{i\in I} A_i = \{x \in \Omega \mid (\forall i \in I)(x \in A_i)\} \tag{7.17}$$

Exercise 7.5.14. Based on Def. 7.5.13, 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 7.5.15 (Equivalence of definitions of independence)**.** Show that our two definitions of independence of k events, Def 7.5.11 and Def 7.5.12, are equivalent.

How many conditions do we need to verify to establish that a list of k events is independent? Definition 7.5.12 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 7.5.16. Show that for $|I| \le 1$, Eq. (7.16) always holds, so in verifying the independence of a list of events, we only need to verify Eq. (7.16) for $|I| \ge 2$. This means verifying $2^k - k - 1$ conditions.

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

Exercise 7.5.17 (Independence of complements-1)**.** Prove: if $A_1, A_2, \dots, A_k$ are independent events then $\overline{A}_1, A_2, \dots, A_k$ are independent events.

Exercise 7.5.18 (Independence of complements-2)**.** For an event A define $A^1 = A$ and $A^{-1} = \overline{A}$. Prove: if $A_1, \dots, A_k$ are independent events and $\epsilon_1, \dots, \epsilon_k \in \{1, -1\}$ then $A_1^{\epsilon_1}, \dots, A_k^{\epsilon_k}$ are independent events.

Exercise 7.5.19. Assume there exist k nontrivial independent events in our probability space. Prove: $|\Omega| \ge 2^k$.

Exercise 7.5.20. For all $n \ge 2$, 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* 7.5.21 (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. 7.5.21 (b).

Exercise* 7.5.22 (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** 7.5.23 (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. 7.5.21 (a) shows that this bound is tight.

Exercise* 7.5.24. 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 7.5.25 (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 7.5.26 (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.

7.6 Random graphs: The Erdős–Rényi model

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

Exercise 7.6.1. $|\Gamma_n| = 2^{\binom{n}{2}}$

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 7.6.2 (Random graphs)**.** Let $A(i,j)$ denote the event that vertices i and j are adjacent ($1 \le i, j \le n$, $i \neq j$). 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 7.6.3. Let $G = (V, E)$ be a graph. The **distance** $\operatorname{dist}(u, v)$ of vertices u and v is the length of a shortest path between them. The **diameter** of G is the maximum distance:

$$\operatorname{diam}(G) = \max_{u,v \in V} \operatorname{dist}(u, v)\,. \tag{7.18}$$

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

Exercise 7.6.4. 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 $\lim_{n\to\infty} p_n = 1$.

7.7 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.

$$n! \sim (n/\mathrm{e})^n\sqrt{2\pi n}. \tag{7.19}$$

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:

$$n! = (n/\mathrm{e})^n\sqrt{2\pi n}(1 + \theta_n/(12n)), \tag{7.20}$$

where $|\theta_n| \le 1$.

Exercise 7.7.1. 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 $\le x$, 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

$$\pi(x) \sim x/\ln x. \tag{7.21}$$

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

$$\sum_{p\le x} 1/p \sim \ln\ln x. \tag{7.22}$$

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

$$\lim_{x\to\infty}\left(\sum_{p\le x} 1/p - \ln\ln x\right) = B. \tag{7.23}$$

(Deduce (7.22) from (7.23).)

Exercise 7.7.2. 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.)

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

Definition 7.8.1. 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 7.8.2. The **expected value** of a random variable X is

$$\mathrm{E}(X) = \sum_{a \in \Omega} X(a) \Pr(a) . \tag{7.24}$$

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

$$\mathrm{E}(X) = \sum_{u \in \mathbb{R}} u \cdot \Pr(X = u) = \sum_{u \in \mathrm{range}(X)} u \cdot \Pr(X = u) . \tag{7.25}$$

Here range(X) denotes the range of X:

$$\mathrm{range}(X) = \{X(a) \mid a \in \Omega\} . \tag{7.26}$$

Exercise 7.8.4. The middle term in Eq. 7.25 seems like an infinite sum. Verify that all but a finite number of terms are zero.

Exercise 7.8.5.

$$\min X \leq \mathrm{E}(X) \leq \max X. \tag{7.27}$$

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 7.8.6 (Additivity of expectation)**.** Let X_1,. . . , X_k be random variables. Then

$$\mathrm{E}(X_1 + \cdots + X_k) = \sum_{i=1}^{k} \mathrm{E}(X_i) \tag{7.28}$$

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

$$\mathrm{E}\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} c_i \mathrm{E}(X_i). \tag{7.29}$$

Definition 7.8.8. 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 7.8.9. 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

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

Exercise 7.8.10 (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 7.8.11. The expected value of an indicator variable is the probability of the event it indicates:

$$\mathrm{E}(\vartheta_A) = \Pr(A)\,. \tag{7.30}$$

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

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 7.8.12. (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 7.8.13. 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 7.8.14 (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 7.8.15 (Poker hand)**.** (a) What is the expected number of Aces in a poker hand? (b) What is the expected number of Spades?

Exercise 7.8.16 (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 7.8.17 (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 7.8.18 (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 7.8.19 (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 7.8.20 (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 7.8.21 (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 7.8.22 (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 7.8.23 (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 $\sim \ln n$.

7.9 Variance, covariance, Chebyshev's Inequality

Definition 7.9.1. 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 7.9.2. 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 7.9.3. Prove: $\mathrm{Var}(X) \geq 0$ and $\mathrm{Var}(X) = 0$ if and only if X is almost constant.

Definition 7.9.4. The **standard deviation** of X is $\sigma(X) := \sqrt{\mathrm{Var}(X)}$.

Exercise 7.9.5. (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 7.9.6. Prove: if c is a constant then $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$; and consequently, $\sigma(cX) = |c|\sigma(X)$.

Exercise 7.9.7. $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

Corollary 7.9.8 (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 7.9.9 (Chebyshev's Inequality)**.** *Let $m = \mathrm{E}(X)$. Then for any number $a > 0$,*

$$\Pr(|X - m| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}. \tag{7.31}$$

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

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \geq \left(\sum_{i=1}^{n} x_i y_i\right)^2. \tag{7.32}$$

Deduce this inequality from Corollary 7.9.8.

Exercise 7.9.11. 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 7.9.12 (Covariance)**.** The **covariance** of the random variables X, Y is defined as

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X) \cdot \mathrm{E}(Y)\,. \tag{7.33}$$

Exercise 7.9.13. $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$

Definition 7.9.14. 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 7.9.15 (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 7.9.16 (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 7.9.17 (Variance of sum)**.** Let $Y = X_1 + \cdots + X_n$ be a sum of random variables. Then

$$\mathrm{Var}(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\cdot \sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j)\,. \tag{7.34}$$

Corollary 7.9.18 (Additivity of variance)**.** *If $X_1, \dots, X_n$ are pairwise uncorrelated random variables then*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)\,. \tag{7.35}$$

In particular, this equation holds if the variables are pairwise independent (see Ex. 7.10.3).

Exercise 7.9.19 (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 7.9.20 (Limit on strongly negatively correlated events)**.** (a) Suppose the events $A_1, \dots, A_m$ are balanced (have probability 1/2) and for each $i \neq j$, $\Pr(|A_i \cap A_j| \le 1/5$. Prove: $m \le 6$. (b) Generalize the statement to events of probability p, with $p^2 - \epsilon$ in the place of 1/5.

7.10 Independence of a pair of random variables

We again fix our probability space.

Definition 7.10.1 (Independence of a pair of random variables)**.** Let X, Y be random variables. We say that X and Y are **independent** if

$$(\forall u, v \in \mathbb{R})(\Pr(X = u \text{ and } Y = v) = \Pr(X = u)\cdot\Pr(Y = v)\,. \tag{7.36}$$

Exercise 7.10.2. If Y is almost constant (see Def. 7.8.1) then X and Y are independent.

Exercise* 7.10.3 (Independent implies uncorrelated)**.** If X, Y are independent then they are uncorrelated, i. e.,

$$\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)\,. \tag{7.37}$$

The next exercise asserts that the converse is false.

Exercise 7.10.4. 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.

7.11 Independence of random variables

Definition 7.11.1. $X_1, \dots, X_k$ are **independent** if

$$(\forall u_1, \dots, u_k \in \mathbb{R}) \left(\Pr\left(X_1 = u_1, \dots, X_k = u_k\right) = \prod_{i=1}^{k} \Pr(X_i = u_i \right)\,. \tag{7.38}$$

Exercise 7.11.2. Prove that the events $A_1, \dots, A_k$ are independent if and only if their indicator variables are independent. – This is less obvious than it seems.

Exercise 7.11.3. Prove that the random variables $X_1, \dots, X_k$ are independent if and only if for all choices of the numbers $u_1, \dots, u_k$, the k events $X_1 = u_1, \dots, X_k = u_k$ are independent. Show that this is also equivalent to the independence of all k-tuples of events of the form $X_1 < u_1, \dots, X_k < u_k$.

Exercise 7.11.4. Prove: if $X_1, \dots, X_k$ are independent then $f_1(X_1), \dots, f_k(X_k)$ 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 7.11.5. 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 7.11.6. Let $X_1, \dots, X_m$ be non-almost-constant random variables (see Def. 7.8.1) 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 7.11.7 (Multiplicativity of the expected value)**.** *If $X_1, \dots, X_m$ are independent, then*

$$\mathrm{E}(\prod_{i=1}^{m} X_i) = \prod_{i=1}^{m} \mathrm{E}(X_i). \tag{7.39}$$

Exercise 7.11.8. Prove this result for indicator variables.

Exercise 7.11.9. 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 7.11.10. Combine the two preceding exercises to a proof of the Theorem for $m = 2$ variables.

Exercise 7.11.11. Deduce the general case from the preceding exercise by induction on m, using Exercise 7.11.5.

This sequence completes the proof of Theorem 7.11.7. □

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. 7.9.18).

Corollary 7.11.12. *Let $X_1, \dots, X_n$ 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 7.11.13 (Weak law of large numbers)**.** *Let $X_1, X_2, \dots$ 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 $\delta > 0$,*

$$\lim_{n\to\infty} \Pr(|Y_n - m| > \delta) = 0. \tag{7.40}$$

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

Remark 7.11.14. 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 $\sigma \ge 0$. 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 $X_{n,1}, X_{n,2}, \dots, X_{n,n}$. Let $Y_n = (1/n)\sum_{i=1}^{n} X_{n,i}$. Then for any $\delta > 0$, the limit relation (7.40) holds.

Exercise 7.11.15. 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.

7.12 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

$$\Pr(|X - n/2| \geq r\sqrt{n}) < \frac{1}{4r^2}. \tag{7.41}$$

Below we shall prove the much stronger inequality

$$\Pr(|X - n/2| \geq r\sqrt{n}) < 2\mathrm{e}^{-2r^2}. \tag{7.42}$$

under the same conditions.

The following corollary illustrates the power of inequality (7.42).

Corollary 7.12.1. *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 $V = \{1, \dots, n\}$ 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 (7.42), we have that

$$\Pr(|\delta_i - (n-1)/2| \geq r\sqrt{n-1}) < 2\mathrm{e}^{-2r^2}. \tag{7.43}$$

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

$$\Pr(|\delta_i - (n-1)/2| \geq \varepsilon(n-1)) < 2\mathrm{e}^{-2\varepsilon^2(n-1)}. \tag{7.44}$$

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 (7.41) will fail. Indeed, setting $r = \varepsilon\sqrt{n-1}$ in inequality (7.41), 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 (7.42). 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 7.12.2 (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 $a > 0$,*

$$\Pr(Y \geq a) < \mathrm{e}^{-a^2/2n} \tag{7.45}$$

and

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

Exercise 7.12.3. Deduce inequality (7.42) 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 7.12.2 to the X_i and translate the result back to X.

Exercise 7.12.4. 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 7.12.5. 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 7.12.2.

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 7.11.4. Therefore we can apply the multiplicativity of the expected value to them to calculate the moment generator function of Y:

$$\mathrm{E}(\mathrm{e}^{tY}) = \mathrm{E}(\exp(\sum_{i=1}^{n} tX_i)) = \mathrm{E}(\prod_{i=1}^{n} Z_i) = \prod_{i=1}^{n} \mathrm{E}(Z_i) = \prod_{i=1}^{n} \mathrm{E}(\exp(tX_i)). \tag{7.47}$$

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

$$\Pr(Y \geq a) = \Pr(\mathrm{e}^{tY} \geq \mathrm{e}^{ta}) \leq \prod_{i=1}^{n} \mathrm{E}(\exp(tX_i))\mathrm{e}^{-ta}. \tag{7.48}$$

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

$$\mathrm{E}(\exp(tX_i)) = \cosh(t). \tag{7.49}$$

Therefore the preceding inequality implies that

$$\Pr(Y \geq a) < \frac{\cosh(t)^n}{\mathrm{e}^{ta}}. \tag{7.50}$$

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 7.12.6. *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

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \dots \tag{7.51}$$

On the right-hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} = 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \dots . \tag{7.52}$$

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 (7.50) we infer that

$$\Pr(Y \geq a) < \exp(t^2 n/2 - ta). \tag{7.53}$$

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 7.12.7 (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 $a > 0$,*

$$\Pr(Y \geq a) < \mathrm{e}^{-a^2/2n} \tag{7.54}$$

and

$$\Pr(|Y| \geq a) < 2\mathrm{e}^{-a^2/2n}. \tag{7.55}$$

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

$$h_t(x) = \cosh(t) + x \cdot \sinh(t). \tag{7.56}$$

(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

$$\mathrm{E}(\mathrm{e}^{tX_i}) \leq \mathrm{E}(h_t(X_i)) = \cosh(t). \tag{7.57}$$

From here on the proof is identical with the proof of Theorem 7.12.2. As before, we set $t = a/n$. □

Exercise 7.12.8. Prove: for almost all graphs $G = (V, E)$ with n vertices,

$$(\forall x \in V)(0.49n < \deg(x) < 0.51n)\,. \tag{7.58}$$

In other words, if p_n denotes the probability of the event described in Eq. (7.58) then $\lim_{n\to\infty} p_n = 1$.

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

Exercise 7.12.9. 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:

$$(\forall x \neq y \in V)(0.24n < N(x, y) < 0.26n)\,. \tag{7.59}$$

7.13 Problems

Exercise 7.13.1. (Bipartite Ramsey) (Erdős) Let $n = 2^{t/2}$, where t is an even integer. Prove that it is possible to color the edges of $K_{n,n}$ red and blue (each edge receives one color) such that there will be no monochromatic $K_{t,t}$. *Hint.* Use the probabilistic method.

A *random graph on n vertices* is defined by fixing a set of n vertices, say $V = [n]$, and flipping a fair coin $\binom{n}{2}$ times to decide adjacency of the $\binom{n}{2}$ pairs of vertices. Let $\mathcal{G}_n$ denote a random graph on the vertex set $[n]$.

Exercise 7.13.2. (Diameter of a random graph)

(a) State the size of the sample space of the experiment which produces a random graph.

(b) What is the probability $\operatorname{diam}(\mathcal{G}_n) = 1$? Your answer should be a very simple closed-form expression. ($\operatorname{diam}(G)$ denotes the diameter of G. See the handout for the definition.)

(c) Prove that almost all graphs have diameter 2.

The meaning of this statement is the following. Let p_n denote the probability that a random graph on n vertices has diameter 2. Then $\lim_{n\to\infty} p_n = 1$.

Hint. Let $q_n = 1 - p_n$. Prove that $q_n \to 0$. Show this by proving that with large probability, every pair of vertices has a common neighbor. What is the probability that vertices x and y do not have a common neighbor? Give a precise answer to this question; it should be a simple formula. Now *estimate* the probability that there exist vertices x, y without a common neighbor.

Use without proof the following fact from calculus:

$$(\forall c, d > 0)(\lim_{x\to\infty} x^c \mathrm{e}^{-dx} = 0).$$

Exercise 7.13.3. (Chromatic number of a random graph) (Erdős) Recall from the graph theory handout that $\omega(G)$ denotes the size of the largest clique (complete subgraph) in the graph G; $\alpha(G)$ denotes the size of the largest independent set (anticlique) in G, and $\chi(G)$ denotes the chromatic number of G. Note that $\alpha(G) = \omega(\overline{G})$ where $\overline{G}$ denotes the complement of G. Note also (do!) that for every graph G, $\chi(G) \ge \omega(G)$.

1. prove: $\chi(G) \ge n/\alpha(G)$, where n is the number of vertices of G.

2. Show that the chromatic number can be *much* greater than the clique number by proving that there exists a constant $c > 0$ such that for all sufficiently large n there exists a graph G_n with n vertices such that

$$\frac{\chi(G_n)}{\omega(G_n)} \ge \frac{cn}{(\log n)^2}.$$

Estimate the value of c in your proof.

Hint. To prove the existence of these graphs, use the probabilistic method. To obtain a lower bound on $\chi(G_n)$, give an upper bound on $\alpha(G_n)$ for almost all graphs G_n.

3. Prove: for almost all graphs, $\chi(G) = \Theta(n/\log n)$. (The lower bound is easy; the upper bound is more challenging!)

Exercise 7.13.4. (Chromatic number of set systems) (Erdős) Let $\mathcal{F} = \{A_1, \dots, A_m\}$ be an r-uniform set-system ($|A_i| = r$) over the universe $[n]$ (so $A_i \subset [n]$). Assume $m \le 2^{r-1}$. Prove that $\mathcal{F}$ is 2-colorable, i. e., it is possible to color every vertex $v \in [n]$ red or blue such that none of the A_i is monochromatic (each A_i has both colors). *Hint.* Assign the colors at random. Compute the expected number of monochromatic sets A_i.

Exercise 7.13.5. (Error-correcting codes) Let X be a set of n elements. Let $\mathcal{B}(X)$ be the set of all subsets of X; we view $\mathcal{B}(X)$ as a uniform probability space. A "random subset of X" is an element of $\mathcal{B}(X)$ chosen from the uniform distribution.

(a) Prove: $E(|A \setminus B|) = n/4$, where A, B are two independent random subsets of X. What is the size of the sample space for this experiment?

(b) (Constant-rate, cn-error-correcting codes) Prove that there exists a constant $C > 1$ and there exists a family $\{A_1, \dots, A_m\}$ of $m \ge C^n$ subsets of X such that $(\forall i, j)(i \ne j \Rightarrow |A_i \setminus A_j| \ge 0.24n)$. *Hint.* Take m random subsets, chosen independently. Use Chernoff's inequality to prove that $|A_i \setminus A_j| < 0.24n$ is exponentially unlikely. *Explanation of the title.* Suppose we want to send messages ((0, 1)-strings) of length k through a noisy channel. Let $n = k/\log C$, so $2^k = C^n = m$ and we can think of the messages as integers from 1 to m. Rather than sending message i, we transmit the incidence vector of the set A_i. This increases the length of the message by a constant factor $(1/\log C)$. On the other hand, even if 23% of the transmitted bits get changed due to noise, the error can uniquely be corrected because the difference (Hamming distance) between any two valid codewords is at least $0.48n$. – Here we only prove the existence of such codes. Constructive versions exist (Justesen codes).

Exercise 7.13.6. (Strongly negatively correlated events) Let $A_1, \dots, A_m$ be events with probability 1/2; suppose $(\forall i, j)(i \ne j \Rightarrow P(A_i \cap A_j) \le 1/5)$. Prove: $m \le 6$. *Hint.* Use the Cauchy– Schwarz inequality, Corollary 7.9.8.

Chapter 8

Finite Markov Chains

Exercises. The unmarked exercises are routine, the exercises marked with a "plus" (+) are creative, those marked with an asterisk (*) are challenging.

Recall that a **directed graph** (digraph, for short), is a pair $G = (V, E)$, where V is the set of "vertices" and E is a set of ordered pairs of vertices called "edges:" $E \subseteq V \times V$.

A *discrete system* is characterized by a set V of "states" and *transitions* between the states. V is referred to as the **state space**. We think of the transitions as occurring at each time beat, so the state of the system at time t is a value $X_t \in V$ $(t = 0, 1, 2, \dots)$. The adjective "discrete" refers to discrete time beats.

A *discrete stochastic process* is a discrete system in which transitions occur randomly according to some probability distribution. The process is *memoryless* if the probability of an $i \to j$ transition does not depend on the history of the process (the sequence of previous states): $(\forall i, j, u_0, \dots, u_{t-1} \in V)(P(X_{t+1} = j \mid X_t = i, X_{t-1} = u_{t-1}, \dots, X_0 = u_0) = P(X_{t+1} = j \mid X_t = i))$. (Here the universal quantifier is limited to feasible sequences of states $u_0, u_1, \dots, u_{t-1}, i$, i. e., to sequences which occur with positive probability; otherwise the conditional probability stated would be undefined.) If in addition the transtion probability $p_{ij} = P(X_{t+1} = j \mid X_t = i\}$ does not depend on the time t, we call the process *homogeneous.*

A **finite Markov chain** is a memoryless homogeneous discrete stochastic process with a finite number of states.

Let $\mathcal{M}$ be a finite Markov chain with n states, $V = [n] = \{1, 2, \dots, n\}$. Let p_{ij} denote the probability of transition from state i to state j, i. e., $p_{ij} = P(X_{t+1} = j \mid X_t = i)$. (Note that this is a conditional probability: the question of $i \to j$ transition only arises if the system is in state i, i. e., $X_t = i$.)

The finite Markov chain $\mathcal{M}$ is characterized by the $n \times n$ **transition matrix** $T = (p_{ij})$ $(i, j \in [n])$ and an **initial distribution** $q = (q_1, \dots, q_n)$ where $q_i = P(X_0 = i)$.

Definition. An $n \times n$ matrix $T = (p_{ij})$ is **stochastic** if its entries are nonnegative real numbers and the sum of each row is 1:

$$(\forall i, j)(p_{ij} \geq 0) \text{ and } (\forall i)(\textstyle\sum_{j=1}^{n} p_{ij} = 1).$$

Exercise 8.1.1. The transition matrix of a finite Markov chain is a stochastic matrix. Conversely, every stochastic matrix can be viewed as the transition matrix of a finite Markov chain.

Exercise 8.1.2. Prove: if T is a stochastic matrix then T^k is a stochastic matrix for every k.

Random walks on digraphs are important examples of finite Markov chains. They are defined by hopping from vertex to neighboring vertex, giving equal chance to each out-neighbor. The state space will be V, the set of vertices. The formal definition follows.

Let $G = (V, E)$ be a finite digraph; let $V = [n]$. Assume $(\forall i \in V)(\deg^+(i) \geq 1)$. Set $p_{ij} = 1/\deg^+(i)$ if $(i, j) \in E$; $p_{ij} = 0$ otherwise.

Exercise 8.1.3. Prove that the matrix (p_{ij}) defined in the preceding paragraph is stochastic.

Conversely, all finite Markov chains can be viewed as *weighted* random walks on a digraph, the weights being the transition probabilities. The formal definition follows.

Let $T = (p_{ij})$ be an arbitrary (not necessarily stochastic) $n \times n$ matrix. We associate with T a digraph $G = (V, E)$ as follows. Let $V = [n]$ and $E = \{(i, j) : p_{ij} \neq 0\}$. We label the edge $i \to j$ with the number $p_{ij} \neq 0$ (the "weight" of the edge).

This definition makes sense for any matrix T; edges indicate nonzero entries. If T is the transition matrix of a finite Markov chain $\mathcal{M}$ then we call the associated digraph the **transition digraph** of $\mathcal{M}$. The **vertices** of the transition digraph represent the **states** of $\mathcal{M}$ and the **edges** the **feasible transitions** (transitions that occur with positive probability).

Exercise 8.1.4. Prove that in the transition digraph of a finite Markov chain, $(\forall i)(\deg^+(i) \geq 1)$.

Exercise 8.1.5. Draw the transition digraph corresponding to the stochastic matrix

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

Label the edges with the transition probabilities.

The principal subject of study in the theory of Markov chains is the **evolution** of the system.

The *initial distribution* $q = (q_1, \ldots, q_n)$ describes the probability that the system is in a particular state at time $t = 0$. So $q_i \geq 0$ and $\sum_{i=1}^{n} q_i = 1$.

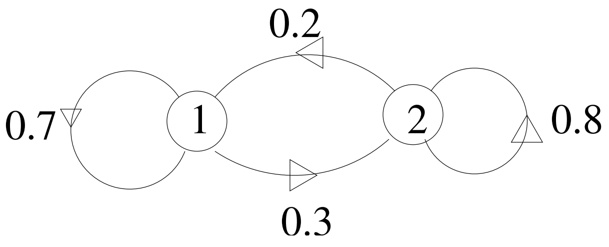


Figure 8.1: A Markov Chain with two states

Set $q(0) = q$ and let $q(t) = (q_{1t}, \dots, q_{nt})$ be the distribution of the states at time t, i. e., the distribution of the random variable X_t:

$$q_{it} = P(X_t = i).$$

The following simple equation describes the evolution of a finite Markov chain.

Exercise 8.1.6. (Evolution of Markov chains) Prove: $q(t) = q(0)T^t$.

So the study of the *evolution of a finite Markov chain* amounts to studying the *powers of the transition matrix.*

Exercise 8.1.7. Experiment: study the powers of the matrix A defined in Exercise 8.1.5. Observe that the sequence $I, A, A^2, A^3, \dots$ appears to converge. What is the limit?

Exercise$^+$ 8.1.8. Prove the convergence observed in the preceding exercise.

The study of the powers rests on the study of *eigenvalues* and *eigenvectors.*

Definition. A **left eigenvector** of an $n \times n$ matrix A is a $1 \times n$ vector $x \neq 0$ such that $xA = \lambda x$ for some (complex) number λ called the *eigenvalue* corresponding to x. A **right eigenvector** of A is an $n \times 1$ matrix $y \neq 0$ such that $Ay = \mu y$ for some (complex) number μ called the *eigenvalue* corresponding to y.

Remember that the zero vector is never an eigenvector.

The right action of a matrix. Note that if $x = (x_1, \dots, x_n)$ is a $1 \times n$ vector, $A = (a_{ij})$ is an $n \times n$ matrix, and $z = (z_1, \dots, z_n) = xA$ then

$$z_j = \sum_{i=1}^{n} x_i a_{ij}. \tag{8.1}$$

Note that if G is the digraph associated with the matrix A then the summation can be reduced to

$$z_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{8.2}$$

So the **left eigenvectors** to the eigenvalue λ is defined by the equation

$$\lambda x_j = \sum_{i:i\to j}^{n} x_i a_{ij}. \tag{8.3}$$

Exercise 8.1.9. State the equations for the left action and the right eigenvectors of the matrix A.

Theorem. The left and the right eigenvalues of a matrix are the same (but not the eigenvectors!).

Proof. Both the right and the left eigenvalues are the roots of the **characteristic polynomial** $f_A(x) = \det(xI - A)$ where I is the $n \times n$ identity matrix.

Exercise 8.1.10. Find the eigenvalues and the corresponding left and right eigenvectors of the matrix A from Exercise 8.1.5.

Hint. The characteristic polynomial is

$$f_A(x) = \begin{vmatrix} x-0.7 & -0.3 \\ -0.2 & x-0.8 \end{vmatrix} = x^2 - 1.5x + 0.5 = (x-1)(x-1/2).$$

So the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1/2$. Each eigenvalue gives rise to a system of linear equations for the coordinates of the corresponding (left/right) eigenvectors.

Exercise$^+$ 8.1.11. Prove: if λ is a (complex) eigenvalue of a stochastic matrix then $|\lambda| \le 1$. *Hint.* Consider a right eigenvector to eigenvalue λ.

Exercise 8.1.12. Let A be an $n \times n$ matrix. Prove: if x is a left eigenvector to eigenvalue λ and y is a right eigenvector to eigenvalue μ and $\lambda \neq \mu$ then x and y are **orthogonal,** i. e., $xy = 0$. *Hint.* Consider the product xAy.

Definition. A **stationary distribution** (also called **equilibrium distribution**) for the Markov chain is a probability distribution $q = (q_1, \dots, q_n)$ ($q_i \ge 0$, $\sum_{i=1}^n q_i = 1$) which is a left eigenvector to the eigenvalue 1: $qA = q$.

Exercise 8.1.13. If at time t, the distribution $q(t)$ is stationary then it will remain the same forever: $q(t) = q(t+1) = q(t+2) = \dots$.

Exercise 8.1.14. Prove: if T is a stochastic matrix then $\lambda = 1$ is a right eigenvalue. *Hint.* Guess the (very simple) eigenvector.

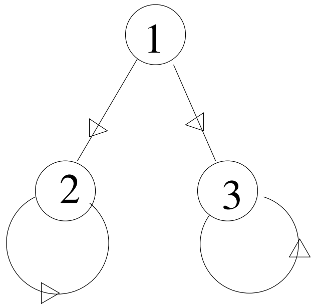


Figure 8.2: A transition digraph

Observe the consequence that $\lambda = 1$ is also a *left* eigenvalue. This is significant because it raises the possibility of having stationary distributions.

Exercise 8.1.15. Find a *left* eigenvector $x = (x_1, x_2)$ to the eigenvalue 1 for the stochastic matrix A defined in Exercise 8.1.5. Normalize your eigenvector such that $|x_1| + |x_2| = 1$. Observe that x is a stationary distribution for A.

Exercise 8.1.16. Let T be a stochastic matrix. Prove: **if** the limit $T^\infty = \lim_{t\to\infty} T^t$ **exists** then every row of T^∞ is a stationary distribution.

Exercise 8.1.17. Consider the stochastic matrix

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Prove that the sequence $I, B, B^2, B^3, \ldots$ does **not** converge, yet B does have a stationary distribution.

Exercise 8.1.18. Let $\vec{C}_n$ denote the directed cycle of length n. Prove that the powers of the transition matrix of the random walk on $\vec{C}_n$ do not converge; but a stationary distribution exists.

Exercise 8.1.19. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$.

Write down the transition matrix of the random walk on the graph shown in Figure 8.2. Prove that the random walk on this graph has 2 stationary distributions.

Definition. A stochastic matrix $T = (p_{ij})$ is called **"doubly stochastic"** if its column sums are equal to 1: $(\forall j \in [n])(\sum_{i=1}^n p_{ij} = 1)$.

In other words, T is doubly stochastic if both T and its transpose are stochastic.

Exercise 8.1.20. Let T be the transition matrix for a finite Markov chain M. Prove that the uniform distribution is stationary if and only if T is doubly stochastic.

A matrix is called **non-negative** if all entries of the matrix are non-negative. The *Perron–Frobenius theory of non-negative matrices* provides the following fundamental result.

Theorem (Perron–Frobenius, abridged) If A is a non-negative $n \times n$ matrix then A has a non-negative left eigenvector.

Exercise 8.1.21. Prove that a non-negative matrix has a non-negative right eigenvector. (Use the Perron–Frobenius Theorem.)

Exercise 8.1.22. Let T be a stochastic matrix and x a non-negative left eigenvector to eigenvalue λ. Prove: $\lambda = 1$. *Hint.* Use Exercise 8.1.12.

Exercise 8.1.23. Prove: **every finite Markov chain has a stationary distribution.**

Exercise$^+$ 8.1.24. Let A be a non-negative matrix, x a non-negative left eigenvector of A, and G the digraph associated with A. Prove: if G is **strongly connected** then all entries of x are **positive.** *Hint.* Use equation (8.3).

Exercise 8.1.25. Let A be a non-negative matrix, x and x' two non-negative eigenvectors of A, and G the digraph associated with A. Prove: if G is **strongly connected** then x and x' belong to the same eigenvalue. *Hint.* Use the preceding exercise and Exercise 8.1.12.

Exercise$^+$ 8.1.26. Let A be a non-negative matrix; let x be a non-negative left eigenvector to the eigenvalue λ and let x' be another left eigenvector with real coordinates to the same eigenvalue. Prove: if G is **strongly connected** then $(\exists \alpha \in \mathbb{R})(x' = \alpha x)$. *Hint.* WLOG (without loss of generality we may assume that) all entries of x are positive (why?). Moreover, WLOG $(\forall i \in V)(x'_i \leq x_i)$ and $(\exists j \in V)(x'_j = x_j)$ (why?). Now prove: if $x_j = x'_j$ and $i \to j$ then $x_i = x'_i$. Use equation (8.3).

Finite Markov chains with a **strongly connected** transition digraph (every state is accessible from every state) are of particular importance. Such Markov chains are called **irreducible.** To emphasize the underlying graph theoretic concept (and reduce the terminology overload), we shall deviate from the accepted usage and use the term **strongly connected Markov chains** instead of the classical and commonly used term "irreducible Markov chains."

Our results are summed up in the following exercise, an immediate consequence of the preceding three exercises.

Exercise 8.1.27. Prove: **A strongly connected finite Markov chain (a) has exactly one stationary distribution; and (b) all probabilities in the stationary distribution are positive.**

As we have seen (which exercise?), strong connectivity is not sufficient for the powers of the transition matrix to converge. One more condition is needed.

Definition. The **period** of a vertex v in the digraph G is the g.c.d. of the lengths of all closed directed walks in G passing through v. If G has no closed directed walks through v, the period of v is said to be 0. If the period of v is 1 then v is said to be **aperiodic.**

Exercise 8.1.28. (a) Show that it is not possible for every state of a finite Markov chain to have period 0 (in the transition digraph). (b) Construct a Markov chain with n states, such that all but one state has period 0.

Note that a **loop** is a closed walk of length 1, so if G has a loop at v then v is automatically aperiodic. A **lazy random walk** on a digraph stops at each vertex with probability 1/2 and divides the remianing 1/2 evenly between the out-neighbors ($p_{ii} = 1/2$, and if $i \to j$ then $p_{ij} = 1/2 \deg +(i)$). So the lazy random walks are aperiodic at each vertex.

Exercise 8.1.29. Let $G = (V, E)$ be a digraph and $x, y \in V$ two vertices of G. Prove: if x and y belong to the same strong component of G (i. e., x and y are mutually accessible from one another) then the periods of x and y are equal.

It follows that **all states of a strongly connected finite Markov chain have the same period.** We call this common value the **period** of the strongly connected Markov chain. A Markov chain is **aperiodic** if every node has period 1.

Exercise 8.1.30. Recall that (undirected) graphs can be viewed as digraphs with each pair of adjacent vertices being connected in both directions. Let G be an undirected graph viewed as a digraph. Prove: every vertex of G has period 1 or 2. The period of a vertex v is 2 if and only the connected component of G containing v is bipartite.

Exercise 8.1.31. Suppose a finite Markov chain $\mathcal{M}$ is strongly connected and NOT aperiodic. (It follows that the period ≥ 2 (why?).)
Prove: the powers of the transition matrix do not converge.
Hint. If the period is d, prove that the transition graph is a "blown-up directed cycle of length d" in the following sense: the vertices of the transition graph can be divided into d disjoint subsets $V_0, V_1, \ldots, V_{d-1}$ such that $(\forall k)$ all edges starting at V_k end in V_{k+1}, where the subscript is read modulo d (wraps around). – Once you have this structure, observe that any t-step transition would take a state in V_k to a state in V_{k+t} (the subscript again modulo d).

Now we state the Perron–Frobenius Theorem in full.

Theorem (Perron–Frobenius, unabridged) Let A be a non-negative $n \times n$ matrix and G the associated digraph. Let $f_A(x) = \prod_{i=1}^{n}(x - \lambda_i)$ be the characteristic polynomial of A factored over the complex numbers. (So the λ_i are the eigenvalues, listed with multiplicity.) Then

(a) There is an eigenvalue λ_1 such that

  (a1) λ_1 is real and non-negative;

  (a2) $(\forall i)(\lambda_1 \geq |\lambda_i|)$;

  (a3) there exists a non-negative eigenvector to eigenvalue λ_1.

(b) If G is strongly connected and **aperiodic** then $(\forall i)(\lambda_1 > |\lambda_i|)$.

Definition. A **strongly connected aperiodic Markov chain** is called **ergodic.**

The significance of aperiodicity is illuminated by the following exercises.

Exercise 8.1.32. Prove that the eigenvalues of the random walk on the directed n-cycle are exactly the n-th roots of unity. (So all of them have unit absolute value.)

More generally, we have the following:

Exercise 8.1.33. Let A be a (not necessarily non-negative) $n \times n$ matrix and G the associated digraph. Suppose d is a common divisor of the periods of G. Let ω be a complex d-th root of unity (i. e., $\omega^d = 1$). Then, if λ is an eigenvalue of A then $\lambda\omega$ is also an eigenvalue of A. *Hint.* Equation (8.3).

The following consequence of the Perron–Frobenius Theorem is the fundamental result in the theory of finite Markov chains.

Exercise* 8.1.34. (Convergence of ergodic Markov chains.) Prove: if T is the transition matrix of an **ergodic Markov chain** then the powers of T **converge.** *Hint.* There exists an invertible complex matrix S such that $U = S^{-1}TS$ is an upper triangular matrix of which the first row is $[1, 0, 0, \dots, 0]$. (This follows, for example, from the Jordan normal form.) Now the diagonal entries of U are the eigenvalues, starting with $\lambda_1 = 1$; all other eigenvalues satisfy $|\lambda_i| < 1$. Prove that as a consequence, the sequence U^t $(t \to \infty)$ converges to the matrix N which has a 1 in the top left corner and 0 everywhere else. Now $T^k \to M := SNS^{-1}$ (why?).

Exercise 8.1.35. Prove: if T is the transition matrix of an ergodic Markov chain and $\lim_{t\to\infty} T^t = M$ then all rows of M are equal.

Exercise 8.1.36. Prove: if a finite Markov chain is ergodic then from any initial distributrion, the process will approach the unique stationary distribution. In other words, let T be the transition matrix, s the stationary distribution, and q an arbitrary initial distribution. Then

$$\lim_{t\to\infty} qT^t = s.$$

The following example illuminates the kind of Markov chains encountered in combinatorics, theoretical computer science, and statistical physics.

Random recoloring: a class of large Markov chains. Let $G = (V, E)$ be a graph with n vertices and maximum degree Δ; and let $Q \geq \Delta + 1$. Let S be the set of all legal colorings of G with Q colors, i. e., S is the set of functions $f : V \to [Q]$ such that if $v, w \in V$ are adjacent then $f(v) \neq f(w)$. This "random recoloring process" is a Markov chain which takes S as its set of states (the "state space"). The transitions from a legal coloring are defined as follows. We pick a vertex $v \in V$ at random, and recolor it by one of the available colors (colors not used by the neighbors of v), giving each available color an equal chance (including the current color of v).

Exercise 8.1.37. Prove: if $Q \geq \Delta + 2$ then the random recoloring process is an ergodic Markov chain.

Exercise 8.1.38. Prove that the number of states of the random recoloring process is between $(Q - \Delta - 1)^n$ and Q^n. So if $Q \geq \Delta + 2$ then the state space is exponentially large.

Exercise 8.1.39. Prove: if $Q \geq \Delta + 2$ then the stationary distribution for the random recoloring process is uniform.

As a consequence, the random recoloring process will converge to a uniformly distributed random legal Q-coloring of G. Just how quickly the process approaches the uniform distribution is an open problem. While the state space is exponential, it is expected that the process distribution will be close to uniform within a polynomial (n^{const}) number of steps. This phenomenon is called **rapid mixing**. Marc Jerrum proved in 1995 that for $Q > 2\Delta$, the random recoloring process does indeed mix rapidly; Jerrum proved an $O(n \log n)$ bound on the mixing time. In a recent (2000) paper, published in the *Journal of Mathematical Physics*, Eric Vigoda showed that the 2Δ bound was not best possible; he proved that rapid mixing already occurs for $Q > (11/6)\Delta$; under this weaker condition Vigoda shows a somewhat less rapid, $O(n^2 \log n)$ mixing. The techniques leading to such improvements are expected to be widely applicable in combinatorics, theoretical computer science, and statistical physics.

Concluding remarks. Markov chains are widely used models in a variety of areas of theoretical and applied mathematics and science, including statistics, operations research, industrial engineering, linguistics, artificial intelligence, demographics, genomics. Markov chain models are used in performance evaluation for computer systems ("if the system goes down, what is the chance it will come back?"), in queuing theory (server queuing, intelligent transportation systems). Hidden Markov models (where the transition probabilities are not known) are a standard tool in the design of intelligent systems, including speech recognition, natural language modelling, pattern recognition, weather prediction.

In discrete mathematics, theoretical computer science, and statistical physics, we often have to consider finite Markov chains with an enormous number of states. Card shuffling is an example of a Markov chain with 52! states. The "random recoloring process," discussed above, is an example of a class of Markov chains which have exponentially many states compared to the length of the description of the Markov chain. (The description of an instance of the random recoloring process consists of specifying the graph G and the parameter Q.) We remark that the random recoloring process is but one instance of a class of Markov chains referred to as "Glauber dynamics," originating in statistical physics.

An example from computer science: if the state of a memory unit on a computer chip can be described by a bit-string of length k then the number of states of the chip is 2^k. (Transitions can be defined by changing one bit at a time.)

This exponential behavior is typical of combinatorially defined Markov chains.

Because of the exponential growth in the number of states, it is not possible to store the transition matrices and to compute their powers; the size of the matrices becomes prohibitive even for moderate values of the description length of the states. (Think of a $52! \times 52!$ matrix to study card shuffling!)

The evolution of such "combinatorially defined" Markov chains is therefore the subject of intense theoretical study. It is of great importance to find conditions under which the distribution is guaranteed to get **close** to the stationary distribution very fast (in a polynomial number of steps). As noted above, this circumstance is called **rapid mixing**. Note that rapid mixing takes place much faster than it would take to visit each state! (Why is this not a paradox?)

8.2 Problems

Exercise 8.2.1. Let $\mathcal{M}$ be the Markov chain shown in Figure 8.3.

1. Is $\mathcal{M}$ strongly connected?
2. Write down the transition matrix T for $\mathcal{M}$.
3. What is the period of vertex 1?
4. Find a stationary distribution for $\mathcal{M}$. You should describe this distribution as a 1×5 matrix.

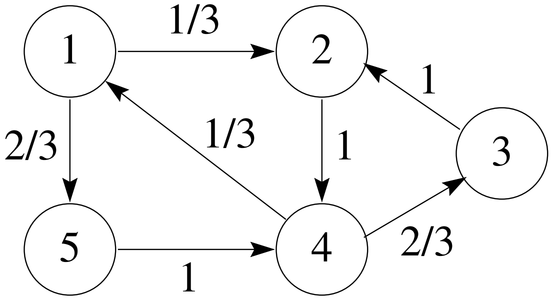


Figure 8.3: Transition graph for a Markov chain.

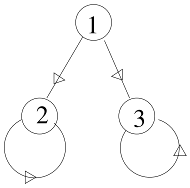


Figure 8.4: The transition graph for a Markov chain.

5. Prove that $\lim_{t\to\infty} T^t$ does not exist. Prove this directly, do not refer to the Perron-Frobenius theorem.

Exercise 8.2.2. Consider the following digraph: $V = [3]$, $E = \{1 \to 2, 1 \to 3, 2 \to 2, 3 \to 3\}$. Write down the transition matrix of the random walk on this graph, with transition probabilities as shown in Figure 8.4. State two different stationary distributions for this Markov chain.