Discrete Mathematics
Lecture Notes
Incomplete Preliminary Version

Instructor: László Babai

Last revision: June 22, 2003
Last update: October 24, 2003

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) $(\forall x)($ if $(\forall y)(x \mid y)$ then $x = \pm 1)$.

(f) $(\forall x)($ if $(\forall y)(y \mid x)$ 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) $(\forall x)(\forall y)(x \equiv y \pmod{0} \iff x = y)$.

(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

2.1 Limit of sequence

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

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

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

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 $c, c, \ldots$) is equivalent to $a_n \to c$ (where c means the number c).

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: $\binom{2n}{n} \sim \dfrac{4^n}{\sqrt{\pi n}}$.

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 $x_1 \leq \cdots \leq x_k$. 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 $(\exists n_0)(\forall n \geq n_0)(a_n \geq 0)$, i. e., $a_n \geq 0$ 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 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 $a_n \sim b_n \Longleftrightarrow a_n = b_n(1 + o(1))$.

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 $\ln a_n = o(\ln b_n)$.

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

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

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

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 $\leq x$. 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 $P(x) \leq 4^x$. For the lower bound, prove that if a prime power p^t divides the binomial coefficient $\binom{n}{k}$ then $p^t \leq n$. From this it follows that $\binom{2n}{n}$ divides the product $P(2n)P((2n)^{1/2})P((2n)^{1/3})P((2n)^{1/4})\ldots$. 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?

Last update: October 24, 2003

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

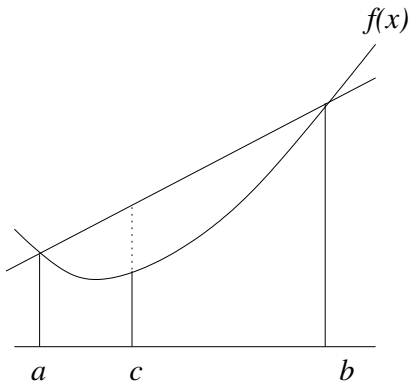
$\lim x^y = 1$ is "almost always true" as $x, y \to 0^+$.

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



$$c = \lambda a + (1-\lambda)b$$

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 $0 \le \lambda \le 1$, the inequality

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

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 $f\left(\dfrac{a+b}{2}\right) \le \dfrac{f(a)+f(b)}{2}$ 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 $x_1, \ldots, x_k$ from the domain of f,*

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

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 $x_1, \ldots, x_k$,

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

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 6.1.22), 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 $x_1, \ldots, x_k > 0$ then

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

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: $a \mid b \iff \operatorname{div}(a) \subseteq \operatorname{div}(b)$.

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

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. $(\forall a)(\forall b \geq 1)(\exists q)(\exists r)(0 \leq r \leq b - 1$ 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: $(\forall a)(\text{g.c.d.}(a, a) = |a|)$.

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

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).

0 Initialize: $A := |a|$, $B := |b|$
1 **while** $B \geq 1$ **do**
2 division: $R := (A \mod B)$
3 $A := B$, $B := R$
4 **end**(**while**)
5 **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 $\leq 2n$ 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 $\{0, 1, \ldots, 2k\}$.

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

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 $n \geq 2$, $F_n = F_{n-1} + F_{n-2}$. 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 $n \geq 1$,

(a) g.c.d. $(F_{n-1}, F_n) = 1$.

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

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

(d) If g.c.d. $(m, n) = d$ then g.c.d. $(a^m - 1, a^n - 1) = a^d - 1$.

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 $(a \bmod m)$ 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 g.c.d. $(a, b) = 0$ 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 $1 \leq a_1 < \cdots < a_{n+1} \leq 2n$ 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 $n \geq 2$. 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 $a, b \in \mathbb{Z}$,

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

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 $\text{g.c.d.}\,(a, b) = 0 \Longleftrightarrow a = b = 0$. Define l.c.m. analogously. When is $\text{l.c.m.}\,(a, b) = 0$?

Exercise 4.2.6. Prove: $\text{g.c.d.}\,(a^k - 1, a^\ell - 1) = a^d - 1$, where $d = \text{g.c.d.}\,(k, \ell)$.

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: $\text{g.c.d.}\,(F_k, F_\ell) = F_d$, where $d = \text{g.c.d.}\,(k, \ell)$.

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

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

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 $e_i = \text{l.c.m.}\,(1, 2, \dots, i)$.

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: $\exists a^{-1} \bmod m \iff \text{g.c.d.}\,(a, m) = 1$.

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 $\text{g.c.d.}\,(a, p) = 1$ 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 $\text{g.c.d.}\,(a, p) = 1$ 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 $\text{g.c.d.}\,(a, m) = 1$ 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 **??**).

Last update: October 24, 2003

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).

$$\begin{aligned}\varphi(n) &= \Big|\big\{k \in [n] \,:\, \text{g.c.d.}\,(k,n) = 1\big\}\Big| \\ &= \text{number of positive integers not greater than } n \text{ which are relatively prime to } n\end{aligned}$$

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 $d_{ij} = \text{g.c.d.}\,(i,j)$. Prove:

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

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

Definition 4.3.5 (Number of [positive] divisors).

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

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).

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

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).

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

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

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. $f\colon \mathbb{N} \to F$ is called **multiplicative** if

$$(\forall a, b)(\text{g.c.d.}\,(a, b) = 1 \Rightarrow f(ab) = f(a)f(b)).$$

f is called **completely multiplicative** if

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

f is called **additive** if

$$(\forall a, b)(\text{g.c.d.}\,(a, b) = 1 \Rightarrow f(ab) = f(a) + f(b)).$$

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 $\lim_{n\to\infty} \frac{p_{n+1}}{p_n} = 1$, 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 $\prod_{\substack{p \le x \\ p \text{ prime}}} p = \exp\big(x(1+o(1))\big)$. Prove that this result is in fact equivalent to the PNT.

Exercise 4.4.10. Let $e_n = \text{l.c.m.}\,(1, 2, \ldots, n)$. 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 $n < p \leq 2n$ 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 $q \leq n$. (*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 $\leq \sqrt{x}$ 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 $p \geq 5$ 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 $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of $\mathbb{R}^n$ $(a_i \in \mathbb{Z})$. The set of *real* linear combinations with $0 \leq a_i \leq 1$ $(a_i \in \mathbb{R})$ form a **fundamental parallelepiped**.

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

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:

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

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

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

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) $\binom{-1}{k}$

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

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.

Last update: October 24, 2003

Exercise 5.1.12. (a) Evaluate the sum $S_n = \sum_{i=0}^{\infty} \binom{n}{i} 2^i$. 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 $a_1, \ldots, a_k$ be positive integers. Prove: the least common multiple $L = \text{l.c.m.}\,(a_1, \ldots, a_k)$ can be expressed through g.c.d's of subsets of the a_i as follows:

$$L = \prod_{I \subseteq [k]} \left(\text{g.c.d.}\,(a_i \,:\, i \in I)\right)^{(-1)^{|I|+1}}.$$

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) $e_n = 1/n$.

Last update: October 24, 2003

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

The graph theoretic terminology we use in class differs from that of Rosen's text. Please remember the differences listed below and use our terminology when it differs from the text's. All concepts 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.

A **graph** (in the 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 vertices. Two vertices joined by an edge are said to be **adjacent**. Two vertices are **neighbors** if they are adjacent. 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 $\binom{V}{2}$ of all 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.

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., $(\forall x, y \in V)(x$ is adjacent o y in $G \Leftrightarrow f(x)$ and $f(y)$ are adjacent in H. Two graphs are *isomorphic* if there *exists* an isomorphism between them.

Exercise 6.1.1. Draw two non-isomorphic regular graphs of the same degree on 6 vertices. Prove that your graphs are not isomorphic.

Exercise 6.1.2. Prove: $\sum_{v \in V} \deg(v) = 2|E|$.

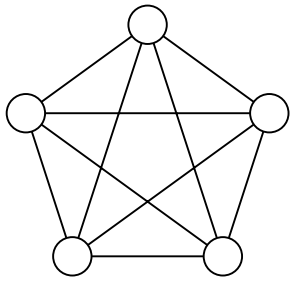


Figure 6.1: The complete graph K_5.

The number of vertices will usually be denoted by n.

Exercise 6.1.3. Observe: $|E| \leq \binom{n}{2}$.

Exercise 6.1.4. Observe: $|E(G)| + |E(\overline{G})| = \binom{n}{2}$.

Exercise 6.1.5. 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 6.1.6. (a) Prove: if $b_n = 2^{\binom{n}{2}}$ and $a_n = b_n/n!$ then $\log_2 a_n \sim \log_2 b_n$.

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

(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.)

Complete graphs, complete bipartite graphs, subgraphs

In a **complete graph**, all pairs of vertices are adjacent. The complete graph on n vertices is denoted by K_n. It has $\binom{n}{2}$ edges. See Figure 6.1.

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). If $k = |V_1|$ and $\ell = |V_2|$ then we obtain the graph $K_{k,\ell}$. This graph has $n = k + \ell$ vertices and $|E| = k\ell$ edges. See Figure 6.2.

The graph $H = (W, F)$ is a **subgraph** of $G = (V, E)$ if $W \subseteq V$ and $F \subseteq E$.

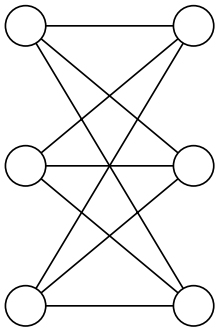


Figure 6.2: The complete bipartite graph $K_{3,3}$.

$H = (W, F)$ is a **spanning subgraph** of G if H is a subgraph and $V = W$.

H is an **induced subgraph** of G if H is a subgraph of G and $(\forall x, y \in W)(x$ and y are adjacent in $H \Leftrightarrow x$ and y are adjacent in G. (So to obtain an induced subgraph, we may delete some vertices and the edges incident with the deleted vertices but no more edges.)

Exercise 6.1.7. Observe: (a) Every graph on n vertices is a spanning subgraph of K_n. (b) All induced subgraphs of a complete graph are complete.

Exercise 6.1.8. Let G be a graph with n vertices and m edges. Count the (a) induced subgraphs of G; (b) the spanning subgraphs of G. Both answers should be very simple expressions.

Exercise 6.1.9. Count those spanning subgraphs of K_n which have exactly m edges.

Exercise 6.1.10. Prove: if G is a bipartite graph with n vertices and m edges then $m \leq \lfloor n^2/4 \rfloor$.

Exercise$^+$ 6.1.11. (Mandel–Turán) Prove: if G is triangle-free ($K_3 \not\subseteq G$) then $|E| \leq \lfloor n^2/4 \rfloor$. Show that this bound is tight for every n. *Hint.* Use induction from in increments of 2; delete both vertices of an edge for the inductive step.

Walks, paths, cycles, trees

This is the area where our terminology most differs from the text.

- **walk** (in text: *path*) of length k: a sequence of $k+1$ vertices $v_0, \ldots, v_k$ such that v_{i-1} and v_i are adjacent for all i.
- **trail** (in text: *simple path*): a walk without repeated edges.

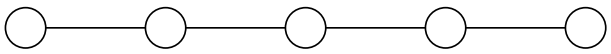


Figure 6.3: P_5, the path of length 4.

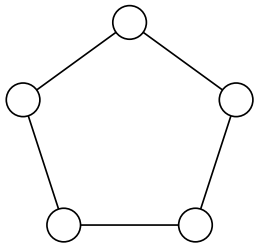


Figure 6.4: C_5, the cycle of length 5.

- **path:** (this all-important concept has no name in the text): a walk without repeated vertices. (Note that the terms "path" and even "simple path" in the text allow vertices to be repeated.) P_{k+1} denotes a path of length k (it has $k+1$ vertices). See Figure 6.3.
- **closed walk** (in text: *circuit* or *cycle*) of length k: a walk $v_0, \ldots, v_k$ where $v_k = v_0$.
- **cycle of length** k or k**-cycle**: (this all-important concept has no name in the text): a closed walk of length k with no repeated vertices except that $v_0 = v_k$. Notation: C_k. See Figure 6.4.
- a graph G is **connected** if there is a path between each pair of vertices.
- a **tree** is a connected graph without cycles. See Figure 6.5.
- H is a **spanning tree** of G if H is a tree and it is a spanning subgraph of G.

Exercise 6.1.12. Prove: if a vertex v has odd degree in the graph G then there exists another vertex w, also of odd degree, such that v and w are connected by a path.

Exercise 6.1.13. 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 6.1.14. Prove that every tree has $n-1$ edges. *Hint.* Induction. Use the preceding exercise.

Exercise 6.1.15. Prove: if G is a connected graph with n vertices then G has at least $n-1$ edges. If G is connedcted and has exactly $n-1$ edges then G is a tree.

Exercise 6.1.16. Draw a copy of each 7-vertex tree. Make sure you don't miss any, and you do not repeat, i. e., no pair of your drawings represent isomorphic trees. State the number of trees you found. Try to list the trees in some systematic fashion.

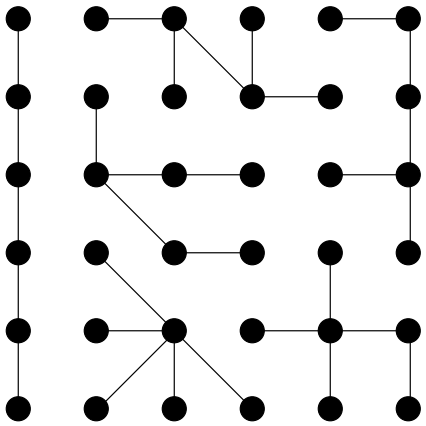


Figure 6.5: The trees on 6 vertices (complete list).

Exercise 6.1.17. Prove: in a tree, all longest paths have a common vertex.

Exercise$^+$ 6.1.18. 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** 6.1.19. (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 6.1.20. 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 6.1.21. 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.* Note that $K_{2,2}$ is isomorphic to C_4, the 4-cycle, therefore $K_{2,2}$ has exactly one 4-cycle. Check also that $K_{2,3}$ has three 4-cycles. Make sure that your answer to the general case conforms with this observation. Your answer should be a very simple formula.

Exercise* 6.1.22. (Kővári–Sós–Turán) Prove: if G has no 4-cycles ($C_4 \not\subseteq G$) then $|E| = O(n^{3/2})$. Show that this bound is tight (apart from the constant implied by the big-Oh notation). *Hint.* Let N denote the number of paths of length 2 in G. Observe that $N = \sum_{i=1}^n \binom{d_i}{2}$ where d_i is the degree of vertex i. On the other hand, observe that $N \le \binom{n}{2}$. (Why? Use the assumption that there are no 4-cycles!) Compare these two expressions for N and apply Jensen's Inequality to the convex function $\binom{x}{2}$.

Cliques, distance, diameter, chromatic number

- A **k-clique** is a subgraph isomorphic to K_k (a set of k pairwise adjacent vertices). $\omega(G)$ denotes the size (number of vertices) of the largest clique in G.
- An **independent set** or **anti-clique** of size k is the complement of a k-clique: k vertices, no two of which are adjacent. $\alpha(G)$ denotes the size of the largest independent set in 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 simple graph is the maximum distance between all pairs of vertices. So if a graph has diameter d then $(\forall x, y \in V)(\text{dist}(x, y) \leq d)$ 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.

  Examples (verify!): trees have infinite girth; the $m \times n$ grid (Figure 6.6) has girth 4 if $m, n \geq 2$; $K_{m,n}$ has girth 4 if $m, n \geq 2$, K_n has girth 3; the Petersen graph (Figure 6.8) has girth 5.
- 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., $\{u, v\} \in E \Rightarrow c(u) \neq c(v)$. A graph is **$k$-colorable** if there exists a legal k-coloring. The **chromatic number** $\chi(G)$ of a graph is the smallest k such that G is k-colorable.
- A graph is **bipartite** if it is 2-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 6.1.23. State the diameter of each of the following graphs: (a) P_n (the path of length $n-1$: this graph has n vertices and $n-1$ edges); (b) C_n (the n-cycle); (c) K_n (the complete graph on n vertices); (d) $K_{n,m}$ (complete bipartite graph);

Exercise 6.1.24. Disprove the following statement: "the diameter of a graph is the length of its longest path." Prove that the statement is true for trees.

Exercise 6.1.25. The $k \times \ell$ grid has $k\ell$ vertices (Figure 6.6). Count its edges.

Exercise 6.1.26. Verify that the diameter of the $k \times \ell$ grid is $k + \ell - 2$.

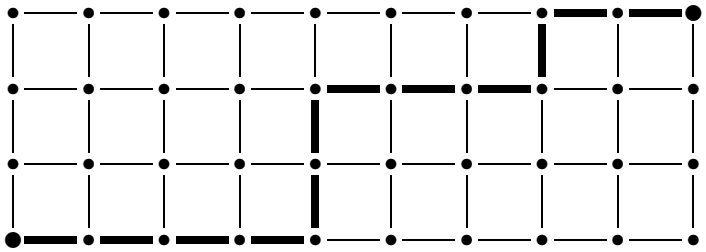


Figure 6.6: The 4×10 grid, with a shortest path between opposite corners highlighted.

Exercise 6.1.27. 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. *Hint.* Think of each shortest path as a sequence of North or East moves, represented as a string over the alphabet $\{N, E\}$.

Exercise 6.1.28. Prove: the bipartite graphs are exactly the subgraphs of the complete bipartite graphs.

Exercise 6.1.29. Prove: a graph is bipartite if and only if it has no odd cycles.

Exercise 6.1.30. 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 6.1.31. 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 6.1.32. 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 6.1.33. 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 6.1.34. 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 6.1.35. (Chromatic number vs. independence number) Prove: if G is a graph with n vertices then $\alpha(G)\chi(G) \geq n$.

Exercise 6.1.36. Give a formal definition of "3-colorable graphs." Watch your quantifiers.

Exercise$^+$ 6.1.37. 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 $2\pi/5$ should not change the picture).

Exercise* 6.1.38. 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 6.1.39. (Erdős, 1959) *Prove: $(\forall k, g)(\exists G)(\chi(G) \geq k$ and G has girth $\geq g$.)*

Exercise 6.1.40. Count the Hamilton cycles in the complete graph K_n.

Exercise 6.1.41. 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 6.1.42. Prove that all grid graphs have a Hamilton path.

Exercise 6.1.43. 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 6.1.44. Prove that the dodecahedron is Hamiltonian. (Lord Hamilton entertained his guests with this puzzle; hence the name.)

Exercise 6.1.45. (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 6.1.46. 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.)

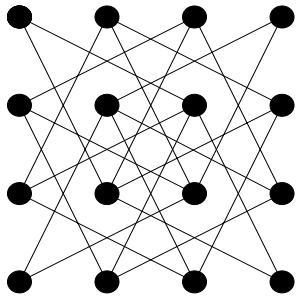


Figure 6.7: Graph of knight moves on a 4×4 chessboard

Exercise 6.1.47. 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 6.1.45, the other inspired by the solution of Exercise 6.1.46.

Exercise 6.1.48. 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 6.1.49. 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 6.1.50. (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 **(Hoffmann–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 or not $r = 57$ can actually occur is open.

Exercise 6.1.51. 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.

Planarity

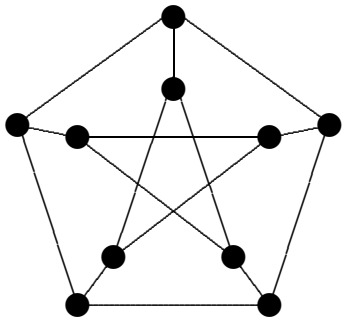


Figure 6.8: The Petersen graph.

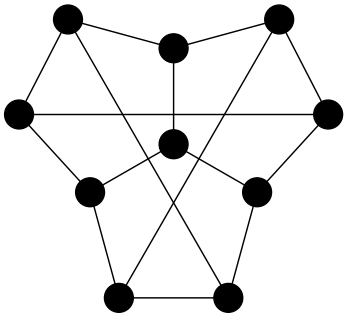


Figure 6.9: Is this graph isomorphic to Petersen's?

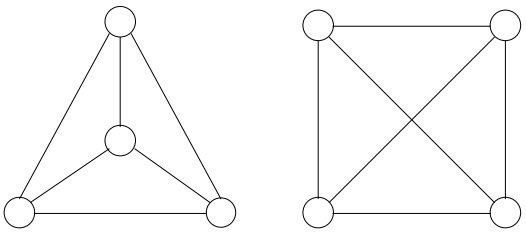


Figure 6.10: K_4 drawn two different ways. Only one is a plane graph.

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).

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 6.1.52. 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 6.1.53. (Jordan's Curve Theorem) *Every plane representation of a cycle has two regions.*

Exercise 6.1.54. (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 6.1.55. Verify that the Platonic solids satisfy Euler's formula.

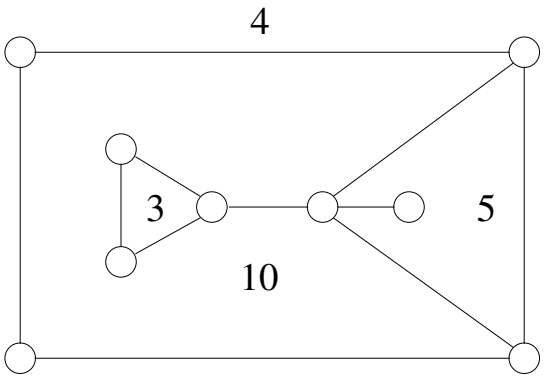


Figure 6.11: The numbers indicate the number of sides of each region of this plane graph.

Exercise 6.1.56. 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 6.1.57. Prove: in a plane graph, $3r \leq 2m$.

Exercise 6.1.58. Prove: in a plane graph without triangles, $2r \leq m$.

Exercise 6.1.59. 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 6.1.60. 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 6.1.61. Prove: the graphs K_5 and $K_{3,3}$ are not planar. *Hint.* Use the preceding two exercises.

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.

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

Theorem 6.1.62. *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 "Kuratowski graphs."

Exercise 6.1.63. 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 6.1.64. 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 6.1.65. Prove that every planar graph has a vertex of degree ≤ 5. *Hint.* $m \leq 3n-6$.

Exercise 6.1.66. 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 6.1.67. 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 6.1.68. Prove that every triangle-free planar graph is 4-colorable.

Ramsey Theory

The Erdős–Rado arrow notation $n \to (k, \ell)$ means 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 6.1.69. Prove: (a) $6 \to (3,3)$; (b) $10 \to (4,3)$; (c) $n \to (n,2)$.

Exercise 6.1.70. (Erdős–Szekeres, 1933)

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

Hint. Induction on $r+s$.

Exercise 6.1.71. Prove: $n \to (k,k)$ where $k = \lceil \log_2 n/2 \rceil$.

Exercise 6.1.72. Define and prove: $17 \to (3,3,3)$.

6.2 Digraph Terminology

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

"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.)

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.**

We shall say that a digraph is **undirected** if the adjacency relation is symmetric ($u \to v$ implies $v \to u$). We say that a digraph is a "graph" if it is undirected and has no *loops*, i. e., no self-adjacencies ($v \not\to v$).

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.

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 for each edge $\{u, v\} \in E_0$, exactly one of (u, v) and (v, u) belongs to E.

Exercise 6.2.2. Suppose the graph G_0 has n vertices and m edges. Count the orientations of G_0.

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 often 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 6.2.3. Count the tournaments on a given set of n vertices. Is the similarity with the number of graphs a coincidence?

Last update: October 24, 2003

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.

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 6.2.4. 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 6.2.5. Prove: if every vertex of a digraph G has the same out-degree d^+ and the same in-degree d^- then $d^+ = d^-$.

An **isomorphism** between the digraphs $G = (V, E)$ and $H = (W, F)$ is a bijection $f : V \rangle W$ 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.

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 6.2.6. List the quadratic residues modulo 5 and modulo 7.

Exercise 6.2.7. Prove that if p is an odd prime then the number of non-congruent quadratic residues modulo p is $(p-1)/2$.

Exercise$^+$ 6.2.8. 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 $V = \{0, 1, \ldots, p-1\}$. Let us set $u \to v$ if $u - v$ is a quadratic residue mod p. $(0 \le u, v \le p-1.)$

Exercise 6.2.9. 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$^+$ 6.2.10. Prove: (a) The Paley tournaments are self-converse. (b) The Paley tournaments are self-complementary.

Exercise* 6.2.11. (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** 6.2.12. (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.

Directed walks, paths, cycles

- **(directed) walk** (in 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 text: *simple path*): a walk without repeated edges.
- **(directed) path:** (this all-important concept has no name in the text): a walk without repeated vertices. (Note that the terms "path" and even "simple path" in the text 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 text: *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 the text): 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 6.2.13. 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, i. e., this relation is *reflexive, symmetric, and transitive.*

- 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.** So the vertices u and v **belong to the same strong component** 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 6.2.14. Prove that a weakly connected digraph has $\geq n-1$ edges; a strongly connected digraph has $\geq n$ edges.

Exercise$^+$ 6.2.15. Prove: if $(\forall v \in V)(\deg^+(v) = \deg^-(v))$ and G is weakly connected then G is strongly connected.

- 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 6.2.16. Prove that every tournament has a Hamilton path.

Exercise$^+$ 6.2.17. Prove that every strongly connected tournament is Hamiltonian.

- A DAG **(directed acyclic graph)** is a digraph with no directed cycles.

Exercise 6.2.18. Prove that for every n, there exists exactly one tournament (up to isomorphism) which is a DAG.

- 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.

For example, if $V = \{1, 2, \ldots, n\}$ and $u \to v$ means $u \neq v$ and $u|v$ (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 6.2.19. Prove that the "divisibility digraph" described in the preceding paragraph has at least $\lfloor n/2 \rfloor !$ topological sorts.

Exercise 6.2.20. Prove that a digraph G can be topologically sorted 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).

The adjacency matrix

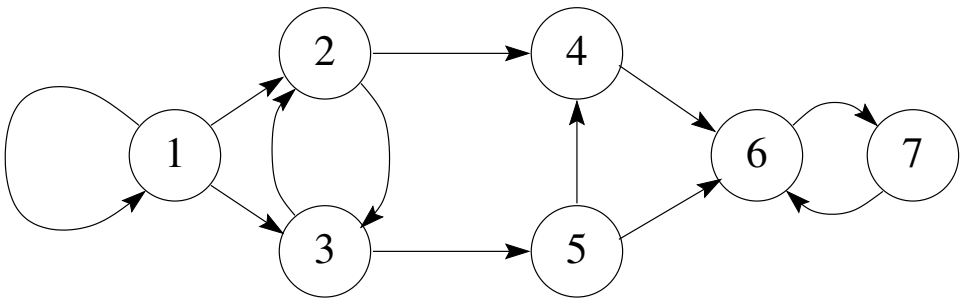
Let $G = (V, E)$ be a digraph; assume $V = [n] = \{1, 2, \ldots, n\}$. Consider the $n \times n$ matrix $A_G = (a_{ij})$ defined as follows: $a_{ij} = 1$ if $i \to j$; and $a_{ij} = 0$ otherwise. A_G is the **adjacency matrix** of G.

Exercise 6.2.21. 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$^+$ 6.2.22. (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 6.2.23. Let T be a tournament with n vertices. Prove: if all vertices have the same out-degree then n is odd.

Exercise 6.2.24. 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 6.2.25. Let $p_1, \ldots, p_k$ 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 $j \,|\, i$ 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 6.2.26. Let v be a vertex in a directed graph. The *period* of v is defined as the g.c.d. of the lengths of all closed walks containing v.

Exercise 6.2.27. 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.

Chapter 7

Finite Probability Spaces

7.1 Finite Probability Spaces and Events

Definition 7.1.1. A **finite probability space** is a finite set $\Omega \neq \emptyset$ together with a function $\Pr : \Omega \to \mathbf{R}^+$ such that

1. $\forall \omega \in \Omega,\ \Pr(\omega) > 0$
2. $\sum_{\omega \in \Omega} \Pr(\omega) = 1.$

The set Ω is the **sample space** and the function Pr is the **probability distribution**. The elements $\omega \in \Omega$ are called **atomic events** or **elementary events**. An **event** is a subset of Ω. For $A \subseteq \Omega$, we define the **probability** of A to be $\Pr(A) := \sum_{\omega \in A} \Pr(\omega)$. In particular, for atomic events we have $\Pr(\{\omega\}) = \Pr(\omega)$; and $\Pr(\emptyset) = 0$, $\Pr(\Omega) = 1$. The **trivial events** are those with probability 0 or 1, i. e. $\emptyset$ and Ω.

The **uniform distribution** over the sample space Ω is defined by setting $\Pr(\omega) = 1/|\Omega|$ for every $\omega \in \Omega$. With this distribution, we shall speak of the **uniform probability space** over Ω. In a uniform space, calculation of probabilities amounts to counting: $\Pr(A) = |A|/|\Omega|$.

Exercise 7.1.2. In the card game of bridge, a deck of 52 cards are 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.

Observation 7.1.3. $\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B)$.

Definition 7.1.4. Events A and B are **disjoint** if $A \cap B = \emptyset$.

Consequence 7.1.5. $\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^{k} \Pr(A_i)$, *and equality holds if and only if the* A_i *are pairwise disjoint.*

Definition 7.1.6. If A and B are events and $\Pr(B) > 0$ then the **conditional probability of** A **relative to** B, written $\Pr(A|B)$, is given by $\Pr(A|B) = \dfrac{\Pr(A \cap B)}{\Pr(B)}$.

We note that B can be viewed as a sample space with the probabilities being the conditional probabilities under condition B.

Note that $\Pr(A \cap B) = \Pr(A|B) \Pr(B)$.

Exercise 7.1.7. Prove: $\Pr(A \cap B \cap C) = \Pr(A|B \cap C) \Pr(B|C) \Pr(C)$.

Exercise 7.1.8. We roll three dice. What is the probability that the sum of the three numbers we obtain is 9? What is the probability that the first die shows 5? What is the conditional probability of this event assuming the sum of the numbers shown is 9? – What is the probability space in this problem? How large is the sample space?

A **partition** of Ω is a family of pairwise disjoint events $H_1, \ldots, H_m$ covering Ω:

$$\Omega = H_1 \cup \ldots \cup H_k, \qquad H_i \cap H_j = \emptyset. \tag{7.1}$$

The sets H_i are the *classes* of the partition. We assume that each class is nonempty.

Exercise 7.1.9. *"Theorem of Complete Probability."* Prove: given a partition $(H_1, \ldots, H_k)$ of Ω, we have

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A|H_i) \Pr(H_i). \tag{7.2}$$

for any event $A \subseteq \Omega$.

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

Definition 7.1.10. Events A and B are **independent** if $\Pr(A \cap B) = \Pr(A) \Pr(B)$.

Exercise 7.1.11. If $\Pr(B) > 0$ then: A and B are independent $\iff \Pr(A|B) = \Pr(A)$.

Exercise 7.1.12. Prove: if A and B are independent events then $\bar{A}$ and B are also independent, where $\bar{A} = \Omega \setminus A$.

Exercise 7.1.13. (a) If we roll a die, are the following events independent: "the number shown is odd"; "the number shown is prime"? (b) 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.

Note that the trivial events are independent of any other events, i. e. if a trivial event is added to a collection of independent events, they remain independent.

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 7.1.14. Are the two events described in Exercise 7.1.8 positively, or negatively correlated, or independent?

Exercise 7.1.15. Prove: two events A, B are positively (negatively) correlated if and only if $\Pr(B|A) > \Pr(B)$ ($\Pr(B|A) < \Pr(B)$, resp.).

Definition 7.1.16. Events $A_1, \ldots, A_k$ are **independent** if for all subsets $S \subseteq \{1, \ldots, k\}$, we have

$$\Pr(\cap_{i \in S} A_i) = \prod_{i \in S} \Pr(A_i). \tag{7.3}$$

Note that if $k \geq 3$, then the statement that events $A_1, \ldots, A_k$ are independent is stronger than pairwise independence. For example, pairwise independence does not imply triplewise independence. For added emphasis, independent events are sometimes called *fully* independent, or *mutually* independent, or *collectionwise* independent.

Exercise 7.1.17. Construct 3 events which are pairwise independent but not collectionwise independent. What is the smallest sample space for this problem?

(See the end of this section for more general problems of this type.)

Exercise 7.1.18. Prove: if the events A, B, C, D, E are independent then the events $A \setminus B$, $C \cup D$, and E are independent as well. Formulate a general statement, for n events grouped into blocks.

Exercise 7.1.19. 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 probaility that the first ball selected is red, assuming condition A. Define p_b and p_g analogously for blue and green outcomes. Assume $p_1 + p_2 = p_3$. Prove: $k \leq 2$. Show that $k = 2$ is actually possible.

Exercise 7.1.20. (Random graphs) Consider the uniform probability space over the set of all the $2^{\binom{n}{2}}$ graphs with a given set V of n vertices. (a) What is the probability that a particular pair of vertices is adjacent? Prove that these $\binom{n}{2}$ events are independent. (b) What is the probability that the degrees of vertex 1 and vertex 2 are equal? Give a closed-form expression. (c) If p_n denotes the probability calculated in part (b), prove that $p_n\sqrt{n}$ tends to a finite positive limit and determine its value. (c) How are the following two events correlated: A:"vertex 1 has degree 3"; B:"vertex 2 has degree 3"? Asymptotically evaluate the ratio $\Pr(A|B)/\Pr(A)$.

In exercises like the last one, one often has to estimate binomial coefficients. The following result comes in handy:

Stirling's formula.

$$n! \sim (n/e)^n\sqrt{2\pi n}. \tag{7.4}$$

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 $a_n \sim f(n)$. Striling'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.

A stronger, non-asymptotic variant, giving useful results for specific values of n, is the following:

$$n! = (n/e)^n\sqrt{2\pi n}(1 + \theta_n/(12n)), \tag{7.5}$$

where $|\theta_n| \le 1$.

Exercise 7.1.21. 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 $\le x$, so $\pi(2) = 1$, $\pi(10) = 4$, etc. The **Prime Number Theorem** of Hadamard and de la Vallée-Poussin asserts that

$$\pi(x) \sim x/\ln x. \tag{7.6}$$

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

$$\sum_{p \le x} 1/p \sim \ln \ln x. \tag{7.7}$$

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

$$\lim_{x \to \infty} \left(\sum_{p \le x} 1/p - \ln \ln x \right) = B. \tag{7.8}$$

(Deduce (7.7) from (7.8).)

Exercise 7.1.22. 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.)

Exercise 7.1.23. Construct a sample space Ω and events $A_1, \ldots, A_n$ ($\forall n \ge 2$) such that $\Pr(A_i) = 1/2$, every $n-1$ of the A_i are independent, but the n events are *not* independent.

Exercise 7.1.24. (*) Let $1 \le k \le n-1$. (a) Construct a sample space Ω and n events such that every k of these n events are independent; but no $k+1$ of these events are independent. (b) Solve part (a) under the additional constraint that each of the n events have probability $1/2$.

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

Exercise 7.1.25. Suppose we have n independent nontrivial events. Prove: $|\Omega| \ge 2^n$.

Exercise 7.1.26. (Small sample space for pairwise independent events.) (a) For $n = 2^k - 1$, construct a probability space of size $n+1$ with n pairwise independent events each of probability $1/2$. (b)* Same for n a prime number of the form $n = 4k - 1$.

Exercise 7.1.27. (*) Prove: if there exist n pairwise independent nontrivial events in a probability space then $|\Omega| \ge n + 1$. (If this is too difficult, solve the special case when all events considered have probability $1/2$ and the space is uniform.)

7.2 Random Variables and Expected Value

Definition 7.2.1. A **random variable** is a function $\xi : \Omega \to \mathbf{R}$.

We say that ξ is **constant** if $\xi(\omega)$ takes the same value for all $\omega \in \Omega$.

Definition 7.2.2. The **expected value** of a random variable ξ is $\mathrm{E}(\xi) = \sum_{\omega\in\Omega} \xi(\omega)\Pr(\omega)$.

Proposition 7.2.3. *Let $\{u_1, \ldots, u_k\}$ be the set of (distinct) values taken by ξ. Let $p_i = \Pr(\xi = u_i)$, where the statement "$\xi = u_i$" refers to the event $\{\omega : \xi(\omega) = u_i\}$. Then $\mathrm{E}(\xi) = \sum_{i=1}^{k} u_i p_i$.*

Proof: Exercise.

Exercise 7.2.4.

$$\min \xi \le \mathrm{E}(\xi) \le \max \xi. \tag{7.9}$$

Throughout these notes, $\xi, \eta, \zeta, \vartheta$, and their subscripted versions refer to random variables.

Proposition 7.2.5. (Additivity of the Expected Value) *Let $\xi_1, \ldots, \xi_k$ be arbitrary random variables. Then*

$$\mathrm{E}(\xi_1 + \cdots + \xi_k) = \sum_{i=1}^{k} \mathrm{E}(\xi_i) \tag{7.10}$$

Proof: $\mathrm{E}(\sum_{i=1}^{k} \xi_i) = \sum_{\omega\in\Omega} (\xi_1(\omega) + \cdots + \xi_k(\omega)) \Pr(\omega) = \sum_{i=1}^{k} \sum_{\omega\in\Omega} \xi_i \Pr(\omega) = \sum_{i=1}^{k} \mathrm{E}(\xi_i).$ □

Exercise 7.2.6. (Linearity of the expected value.) If $c_1, \ldots, c_k$ are constants then

$$\mathrm{E}(\sum_{i=1}^{k} c_i \xi_i) = \sum_{i=1}^{k} c_i \mathrm{E}(\xi_i). \tag{7.11}$$

Definition 7.2.7. The **indicator variable** of an event $A \subseteq \Omega$ is the function $\vartheta_A : \Omega \to \{0, 1\}$ given by

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

Exercise 7.2.8. The expected value of an indicator variable is $\mathrm{E}(\vartheta_A) = \Pr(A)$.

Indicator variables are particularly useful if we want to count events. Some of the exercises at the end of this section should serve as examples.

Exercise 7.2.9. (a) Every random variable ξ is a linear combination of indicator variables. (b) Given a random variable ξ there exist functions $f_1, \ldots, f_k$ such that the random variables $\xi_i := f_i(\xi)$ are indicator variables and ξ is a linear combination of the ξ_i.

We say that ξ is **nonnegative** if $\xi(\omega) \geq 0$ for all $\omega \in \Omega$.

Theorem 7.2.10 (Markov's Inequality). *If ξ is nonnegative then $\forall a > 0$,*

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

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

Exercise 7.2.11. 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 7.2.12. 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 7.2.13. A club with 2000 members distributes membership cards numbered 1 through 2000 to its members at random; each of the 2000! permutations of the cards is equally likely. Members whose card number happens to coincide with their year of birth receive a prize. Determine the expected number of lucky members.

Exercise 7.2.14. What is the expected number of edges in a random graph? What is the expected number of triangles? (There are n vertices; each pair is adjacent with probability 1/2 independently.)

Exercise 7.2.15. 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 7.2.16. 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. Prove your answer. State the size of the sample space for this problem.

Exercise 7.2.17. For a permutation $\pi \in S_n$, let $c_k(\pi)$ denote the number of k-cycles in the cycle decomposition of π. (For instance, if $n = 7$ and $\pi = (13)(256)(47)$ then $c_2(\pi) = 2$, $c_3(\pi) = 1$, and $c_k(\pi) = 0$ for all $k \neq 2, 3$.) Pick π at random (from S_n). Calculate $E(c_k(\pi))$. Your answer should be a very simple expression (no factorials, no binomial coefficients, no summation). Prove your answer.

7.3 Standard deviation and Chebyshev's Inequality

Definition 7.3.1. The k^{th} **moment** of ξ is $\mathrm{E}(\xi^k)$. The k^{th} **central moment** of ξ is the k^{th} moment of $\xi - \mathrm{E}(\xi)$, i. e. $\mathrm{E}((\xi - \mathrm{E}(\xi))^k)$.

Definition 7.3.2. The **variance** of ξ is its second central moment, $\mathrm{Var}(\xi) := \mathrm{E}((\xi - \mathrm{E}(\xi))^2)$.

Note that the variance is always nonnegative. It is zero exactly if ξ is constant. (Why?)

Definition 7.3.3. The **standard deviation** of ξ is $\sigma(\xi) := \sqrt{\mathrm{Var}(\xi)}$.

Exercise 7.3.4. (Invariance under shifts.) Prove that if c is a constant then $\mathrm{Var}(\xi) = \mathrm{Var}(\xi + c)$; and consequently, $\sigma(\xi) = \sigma(\xi + c)$.

Exercise 7.3.5. Prove: if c is a constant then $\mathrm{Var}(c\xi) = c^2 \mathrm{Var}(\xi)$; and consequently, $\sigma(c\xi) = |c|\sigma(\xi)$.

Observation 7.3.6. $\mathrm{Var}(\xi) = \mathrm{E}(\xi^2) - (\mathrm{E}(\xi))^2$.

Corollary 7.3.7 (Cauchy-Schwarz inequality). $(\mathrm{E}(\xi))^2 \leq \mathrm{E}(\xi^2)$. □

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

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

Theorem 7.3.8 (Chebyshev's Inequality). *Let $m = \mathrm{E}(\xi)$. Then for any number $a > 0$,*

$$\Pr(|\xi - m| \geq a) \leq \frac{\mathrm{Var}(\xi)}{a^2}. \tag{7.12}$$

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

Exercise 7.3.9. 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:

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

In other words, if p_n denotes the probability of the event described by the displayed formula then $\lim_{n\to\infty} p_n = 1$.

Exercise 7.3.10. In its more common form the Cauchy-Schwarz inequality asserts that for any real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$ we have

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \geq \left(\sum_{i=1}^{n} x_i y_i\right)^2. \tag{7.13}$$

Deduce this inequality from Corollary 7.3.7.

Exercise 7.3.11. (Limit on negatively correlated events.) Suppose the events $A_1, \ldots, A_m$ each have probability $1/2$ and for each i, j, $\Pr(|A_i \cap A_j| \leq 1/5$. Prove: $m \leq 6$. Generalize the statement to events of probability p, with $p^2 - \epsilon$ in the place of $1/5$.

Exercise 7.3.12. Prove: if the k^{th} moment of ξ is zero for all odd integers $k > 0$ then $\Pr(\xi = u) = \Pr(\xi = -u)$ for all $u \in \mathbf{R}$.

7.4 Independence of random variables

Definition 7.4.1. $\xi_1, \ldots, \xi_k$ are **independent** if $\forall u_1, \ldots, u_k$,

$$\Pr(\xi_1 = u_1, \ldots, \xi_k = u_k) = \prod_{i=1}^{k} \Pr(\xi_i = u_i). \tag{7.14}$$

Exercise 7.4.2. Prove that the events $A_1, \ldots, A_k$ are independent if and only if their indicator variables are independent.

Exercise 7.4.3. Prove that the random variables $\xi_1, \ldots, \xi_k$ are independent if and only if for all choices of the numbers $u_1, \ldots, u_k$, the k events $\xi_1 = u_1, \ldots, \xi_k = u_k$ are independent. Show that this is also equivalent to the independence of all k-tuples of events of the form $\xi_1 < u_1, \ldots, \xi_k < u_k$.

Exercise 7.4.4. Prove: if $\xi_1, \ldots, \xi_k$ are independent then $f_1(\xi_1), \ldots, f_k(\xi_k)$ are also independent, where the f_i are arbitrary functions. For example, ξ_1^2, e^{ξ_2}, and $\cos(\xi_3)$ are independent.

Exercise 7.4.5. Prove: if ξ, η, ζ are independent random variables then $f(\xi, \eta)$ and ζ are also independent, where f is an arbitrary function. (For instance, $\xi + \eta$ and ζ, or $\xi\eta$ and ζ are independent.) Generalize this statement to several variables, grouped into blocks, and a function applied to each block.

Exercise 7.4.6. Let $\xi_1, \ldots, \xi_m$ be non-constant random variables over a sample space of size n. Suppose the ξ_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 $\xi_i\xi_j$ $(1 \leq i < j \leq m)$ 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 $(\forall i)(E(\xi_i) = 0)$; then use the "inner product" argument, using the function $E(\zeta\eta)$ in the role of an "inner product" of the random variables ζ and η.

Theorem 7.4.7 (Multiplicativity of the expected value). *If $\xi_1, \ldots, \xi_m$ are independent, then*

$$\mathrm{E}(\prod_{i=1}^{m} \xi_i) = \prod_{i=1}^{m} \mathrm{E}(\xi_i). \tag{7.15}$$

Exercise 7.4.8. Prove this result for indicator variables.

Exercise 7.4.9. Prove: if ξ, η are independent, then one can write ξ as a sum $\xi = c_1\xi_1 + \ldots + c_k\xi_k$ and η as $\eta = d_1\eta_1 + \ldots + d_\ell\eta_\ell$ where the ξ_i and η_j are indicator variables and for every i, j, the variables ξ_i and η_j are independent.

Exercise 7.4.10. Combine the two preceding exercises to a proof of the Theorem for $m = 2$ variables.

Exercise 7.4.11. Deduce the general case from the preceding exercise by induction on m, using Exercise 7.4.5.

This sequence completes the proof of Theorem 7.4.7. □

While this result required the full force of independence of our random variables, in the next result, only pairwise independence is required.

Theorem 7.4.12 (Additivity of the variance). *Let* $\eta = \xi_1 + \xi_2 + \cdots + \xi_k$. *If* $\xi_1, \ldots, \xi_k$ *are* pairwise independent *then* $\mathrm{Var}(\eta) = \sum_{i=1}^{k} \mathrm{Var}(\xi_i)$.

Proof: By Exercise 7.3.4, we may assume that $\mathrm{E}(\xi_i) = 0$ (otherwise we replace each ξ_i by $\xi_i - \mathrm{E}(\xi_i)$; this will not change the variance, nor does it affect independence (why?)). Having made this assumption it follows that $\mathrm{E}(\eta) = 0$. Moreover, for $i \neq j$ we have $\mathrm{E}(\xi_i \xi_j) = \mathrm{E}(\xi_i)\mathrm{E}(\xi_j) = 0$ by pairwise independence.

It follows that $\mathrm{Var}(\xi_i) = \mathrm{E}(\xi_i^2)$ and $\mathrm{Var}(\eta) = \mathrm{E}(\eta^2) = \mathrm{E}((\sum \xi_i)^2) = \mathrm{E}(\sum_i \xi_i^2 + 2\sum_{i<j} \xi_i \xi_j) = \sum_i \mathrm{E}(\xi_i^2) + 2\sum_{i<j} \mathrm{E}(\xi_i \xi_j) = \sum_i \mathrm{Var}(\xi_i)$. □

Corollary 7.4.13. *Let* $\xi_1, \ldots, \xi_n$ *be random variables with the same standard deviation* σ. *Let us consider their average,* $\eta := (1/n)\sum_{i=1}^{n} \xi_i$. *If the* ξ_i *are pairwise independent then* $\sigma(\eta) = \sigma/\sqrt{n}$. □

Corollary 7.4.14 (Weak law of large numbers). *Let* $\xi_1, \xi_2, \ldots$ *be an infinite sequence of pairwise independent random variables each with expected value* m *and standard deviation* σ. *Let* $\eta_n = (1/n)\sum_{i=1}^{n} \xi_i$. *Then for any* $\delta > 0$,

$$\lim_{n\to\infty} \Pr(|\eta_n - m| > \delta) = 0. \tag{7.16}$$

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

Remark 7.4.15. Strictly speaking, we bent our rules here. An infinite sequence of non-constant, pairwise independent variables requires an infinite sample space. What we actually proved, then, is the following. Let us fix the values m and $\sigma \geq 0$. Assume that we are given an infinite sequence of finite probability spaces, and over the n^{th} space, we are given n independent random variables $\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,n}$. Let $\eta_n = (1/n)\sum_{i=1}^{n} \xi_{n,i}$. Then for any $\delta > 0$, the limit relation (7.16) holds.

Exercise 7.4.16. You and the bank play the following game: you flip n coins: if ξ of them come up "Heads," you receive 2^ξ 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^ξ. 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^ξ. *Hint.* This law talks about the probability that 2^ξ is not within $(1 \pm \epsilon)$-times its expectation.) Prove that the Law does NOT hold for this variable.

7.5 Chernoff's Bound

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 due to Chernoff. The bounds will go to zero exponentially 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 ψ denote the number of heads in this sequence. Then $\mathrm{E}(\psi) = n/2$ and $\mathrm{Var}(\xi) = n/4$ (by the additivity of the variance). Therefore Chebyshev's inequality tells us that

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < \frac{1}{4r^2}. \tag{7.17}$$

Below we shall prove the much stronger inequality

$$\Pr(|\psi - n/2| \geq r\sqrt{n}) < 2e^{-2r^2}. \tag{7.18}$$

under the same conditions.

The following corollary illustrates the power of inequality (7.18).

Corollary 7.5.1. *For any $\varepsilon > 0$, almost all graphs have no vertices of degree $< (1-\varepsilon)n/2$ or $> (1+\varepsilon)n/2$ where n is the number of vertices.*

Proof of the Corollary. Let $V = \{1, \ldots, n\}$ be the vertex set of our random graph. Let δ_i denote the degree of vertex i; so δ_i is the number of heads in a sequence of $(n-1)$ independent coin flips. Therefore, by inequality (7.18), we have that

$$\Pr(|\delta_i - (n-1)/2| \geq r\sqrt{n-1}) < 2e^{-2r^2}. \tag{7.19}$$

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

$$\Pr(|\delta_i - (n-1)/2| \geq \varepsilon(n-1)) < 2e^{-2\varepsilon^2(n-1)}. \tag{7.20}$$

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 $2ne^{-2\varepsilon^2(n-1)}$. 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 (7.17) will fail. Indeed, setting $r = \varepsilon\sqrt{n-1}$ in inequality (7.17), 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 (7.18). It will be convenient to state the main result in terms of random variables with zero expected value.

Theorem 7.5.2 (Chernoff). *Let ξ_i be independent random variables satisfying* $\Pr(\xi_i = 1) = \Pr(\xi_i = -1) = 1/2$. *Let* $\eta = \sum_{i=1}^{n} \xi_i$. *Then for any* $a > 0$,

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{7.21}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{7.22}$$

Exercise 7.5.3. Deduce inequality (7.18) from this theorem.

Hint. Represent ψ as $\sum_{i=1}^{n} \theta_i$ where θ_i is the indicator variable of the i-th coin flip. Set $\xi_i = 2\theta_i - 1$ and $\eta = \sum_{i=1}^{n} \xi_i$. Note that $\psi - n/2 = \eta/2$. Apply Theorem 7.5.2 to the ξ_i and translate the result back to ψ.

Exercise 7.5.4. 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).

Now we turn to the proof of Theorem 7.5.2.

Let t be a positive real number. We shall later suitably choose the value of t. Let us consider the random variables $\zeta_i := \exp(t\xi_i)$. (Notation: $\exp(x) = e^x$.) The ζ_i are again independent (for any fixed t) by Exercise 7.4.4. Therefore we can apply the multiplicativity of the expected value to them:

$$\mathrm{E}(e^{t\eta}) = \mathrm{E}(\exp(\sum_{i=1}^{n} t\xi_i)) = \mathrm{E}(\prod_{i=1}^{n} \zeta_i) == \prod_{i=1}^{n} \mathrm{E}(\zeta_i) = \prod_{i=1}^{n} \mathrm{E}(\exp(t\xi_i)). \tag{7.23}$$

Applying Markov's inequality to the variable $e^{t\eta}$, we conclude that

$$\Pr(\eta \geq a) = \Pr(e^{t\eta} \geq e^{ta}) \leq \prod_{i=1}^{n} \mathrm{E}(\exp(t\xi_i))e^{-ta}. \tag{7.24}$$

Recall that $\cosh(x) = (e^x + e^{-x})/2$ and observe that

$$\mathrm{E}(\exp(t\xi_i)) = \cosh(t). \tag{7.25}$$

Therefore the preceding inequality implies that

$$\Pr(\eta \geq a) < \frac{\cosh(t)^n}{e^{ta}}. \tag{7.26}$$

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 7.5.5. *For all real numbers x,*

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

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

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2} + \frac{x^4}{24} + \frac{x^6}{720} + \dots \tag{7.27}$$

On the right hand side we have

$$\sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} 1 + \frac{x^2}{2} + \frac{x^4}{8} + \frac{x^6}{48} + \dots . \tag{7.28}$$

□

Consequently, from inequality (7.26) we infer that

$$\Pr(\eta \geq a) < \exp(t^2 n/2 - ta). \tag{7.29}$$

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

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

We note that Chernoff's 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 7.5.6 (Chernoff). *Let ξ_i be independent random variables satisfying $|\xi_i| \leq 1$ and $\mathrm{E}(\xi_i) = 0$. Let $\eta = \sum_{i=1}^{n} \xi_i$. Then for any $a > 0$,*

$$\Pr(\eta \geq a) < e^{-a^2/2n} \tag{7.30}$$

and

$$\Pr(|\eta| \geq a) < 2e^{-a^2/2n}. \tag{7.31}$$

Proof: As before, we set $t = a/n$. Let

$$h(x) = \cosh(t) + x \cdot \sinh(t). \tag{7.32}$$

(Recall that $\sinh(t) = (e^t - e^{-t})/2$.) Observe that $h(x) \geq e^{tx}$ for all x in the interval $-1 \leq x \leq 1$. (The graph of $h(x)$ 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 $h(x)$ function, we have $\mathrm{E}(h(\xi_i)) = h(E(\xi_i)) = h(0) = \cosh(t)$. Therefore

$$\mathrm{E}(e^{t\xi_i}) \leq \mathrm{E}(h(\xi_i)) = \cosh(t). \tag{7.33}$$

From here on the proof is identical with the proof of Theorem 7.5.2. □

7.6 Problems

Exercise 7.6.1. (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 7.6.2. (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 $\mathrm{diam}(\mathcal{G}_n) = 1$? Your answer should be a very simple closed-form expression. ($\mathrm{diam}(G)$ 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 7.6.3. (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, $\chi(G) \geq \omega(G)$.

1. prove: $\chi(G) \geq n/\alpha(G)$, 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
$$\frac{\chi(G_n)}{\omega(G_n)} \geq \frac{cn}{(\log n)^2}.$$

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 7.6.4. (Chromatic number of set systems) (Erdős) Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be an r-uniform set-system ($|A_i| = r$) over the universe $[n]$ (so $A_i \subset [n]$). Assume $m \leq 2^{r-1}$. 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 7.6.5. (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 $\{A_1, \ldots, A_m\}$ of $m \geq C^n$ subsets of X such that $(\forall i, j)(i \neq j \Rightarrow |A_i \setminus A_j| \geq 0.24n)$. *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 7.6.6. (Strongly negatively correlated events) Let $A_1, \ldots, A_m$ be events with probability 1/2; suppose $(\forall i, j)(i \neq j \Rightarrow P(A_i \cap A_j) \leq 1/5)$. Prove: $m \leq 6$. *Hint.* Use the Cauchy– Schwarz inequality, Corollary 7.3.7.

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): $(\forall i, j, u_0, \dots, u_{t-1} \in V)(P(X_{t+1} = j \,|\, X_t = i, X_{t-1} = u_{t-1}, \dots, X_0 = u_0) = P(X_{t+1} = j \,|\, X_t = i))$. (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 $p_{ij} = P(X_{t+1} = j \,|\, X_t = i\}$ 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., $p_{ij} = P(X_{t+1} = j \,|\, X_t = i)$. (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.

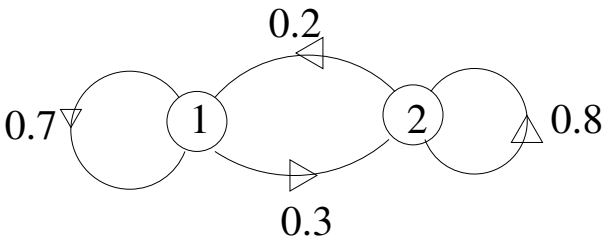


Figure 8.1: NEED CAPTION! AND REF.

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 $\sum_{i=1}^{n} q_i = 1$.

Set $q(0) = q$ and let $q(t) = (q_{1t}, \ldots, q_{nt})$ 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 $I, A, A^2, A^3, \ldots$ 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 $x = (x_1, \ldots, x_n)$ is a $1 \times n$ vector, $A = (a_{ij})$ is an $n \times n$ matrix, and $z = (z_1, \ldots, z_n) = xA$ 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 $\lambda \ne \mu$ 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: $q(t) = q(t+1) = q(t+2) = \ldots$.

Exercise 8.1.14. Prove: if T is a stochastic matrix then $\lambda = 1$ is a right eigenvalue. *Hint.* Guess the (very simple) eigenvector.

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. 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.

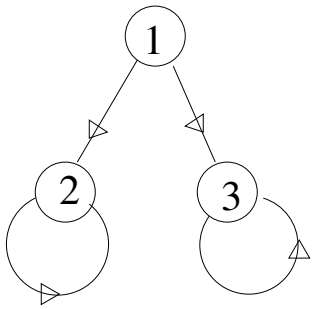


Figure 8.2: A graph with transition probabilities. FIX THIS!

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 $[1, 0, 0, \ldots, 0]$. (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

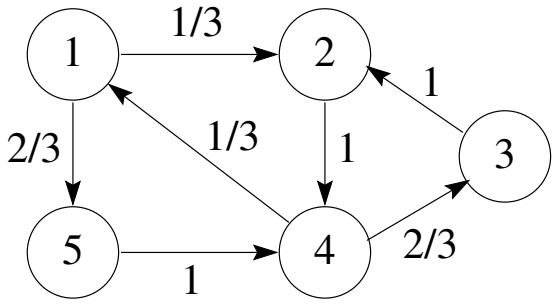


Figure 8.3: Transition graph for a Markov chain.

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 8.2.

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.
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 8.2. State two different stationary distributions for this Markov chain.

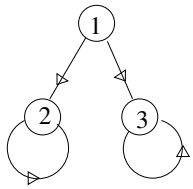


Figure 8.4: The transition graph for a Markov chain.