Wiley Series in Discrete Mathematics and Optimization

Introduction to Combinatorics

SECOND EDITION

MARTIN J. ERICKSON

WILEY

Contents

Introduction to Combinatorics

WILEY SERIES IN DISCRETE MATHEMATICS AND OPTIMIZATION

AARTS AND KORST • Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing

AARTS AND LENSTRA • Local Search in Combinatorial Optimization

ALON AND SPENCER • The Probabilistic Method, Third Edition

ANDERSON AND NASH • Linear Programming in Infinite-Dimensional Spaces: Theory and Application

ARLINGHAUS, ARLINGHAUS, AND HARARY • Graph Theory and Geography: An Interactive View E-Book

AZENCOTT • Simulated Annealing: Parallelization Techniques

BARTHÉLEMY AND GUÉNOCHE • Trees and Proximity Representations

BAZARRA, JARVIS, AND SHERALI • Linear Programming and Network Flows

BRUEN AND FORCINITO • Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century

CHANDRU AND HOOKER • Optimization Methods for Logical Inference

CHONG AND ZAK • An Introduction to Optimization, Fourth Edition

COFFMAN AND LUEKER • Probabilistic Analysis of Packing and Partitioning Algorithms

COOK, CUNNINGHAM, PULLEYBLANK, AND SCHRIJVER • Combinatorial Optimization

DASKIN • Network and Discrete Location: Modes, Algorithms and Applications

DINITZ AND STINSON • Contemporary Design Theory: A Collection of Surveys

DU AND KO • Theory of Computational Complexity

ERICKSON • Introduction to Combinatorics, Second Edition

GLOVER, KLINGHAM, AND PHILLIPS • Network Models in Optimization and Their Practical Problems

GOLSHTEIN AND TRETYAKOV • Modified Lagrangians and Monotone Maps in Optimization

GONDRAN AND MINOUX • Graphs and Algorithms (*Translated by S. Vajdā*)

GRAHAM, ROTHSCHILD, AND SPENCER • Ramsey Theory, Second Edition

GROSS AND TUCKER • Topological Graph Theory

HALL • Combinatorial Theory, Second Edition

HOOKER • Logic-Based Methods for Optimization: Combining Optimization and Constraint Satisfaction

IMRICH AND KLAVŽAR • Product Graphs: Structure and Recognition

JANSON, LUCZAK, AND RUCINSKI • Random Graphs

JENSEN AND TOFT • Graph Coloring Problems

KAPLAN • Maxima and Minima with Applications: Practical Optimization and Duality

LAWLER, LENSTRA, RINNOOY KAN, AND SHMOYS, Editors • The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization

LAYWINE AND MULLEN • Discrete Mathematics Using Latin Squares

LEVITIN • Perturbation Theory in Mathematical Programming Applications
MAHMOUD • Evolution of Random Search Trees
MAHMOUD • Sorting: A Distribution Theory
MARTELLI • Introduction to Discrete Dynamical Systems and Chaos
MARTELLO AND TOTH • Knapsack Problems: Algorithms and Computer Implementations
McALOON AND TRETKOFF • Optimization and Computational Logic
MERRIS • Combinatorics, Second Edition
MERRIS • Graph Theory
MINC • Nonnegative Matrices
MINOUX • Mathematical Programming: Theory and Algorithms (*Translated by S. Vajd*ā)
MIRCHANDANI AND FRANCIS, Editors • Discrete Location Theory
NEMHAUSER AND WOLSEY • Integer and Combinatorial Optimization
NEMIROVSKY AND YUDIN • Problem Complexity and Method Efficiency in Optimization (*Translated by E. R. Dawson*)
PACH AND AGARWAL • Combinatorial Geometry
PLESS • Introduction to the Theory of Error-Correcting Codes, Third Edition
ROOS AND VIAL • Ph. Theory and Algorithms for Linear Optimization: An Interior Point Approach
SCHEINERMAN AND ULLMAN • Fractional Graph Theory: A Rational Approach to the Theory of Graphs
SCHIFF • Cellular Automata: A Discrete View of the World
SCHRIJVER • Theory of Linear and Integer Programming
SPALL • Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control
STIEBITZ, SCHEIDE, TOFT, AND FAVRHOLDT • Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture
SZPANKOWSKI • Average Case Analysis of Algorithms on Sequences
TOMESCU • Problems in Combinatorics and Graph Theory (*Translated by R. A. Melter*)
TUCKER • Applied Combinatorics, Second Edition
WOLSEY • Integer Programming
YE • Interior Point Algorithms: Theory and Analysis

Introduction to Combinatorics

Second Edition

Martin J. Erickson
Department of Mathematics
Truman State University
Kirksville, MO

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data is now available.

ISBN 978-1-118-63753-1

To my parents, Robert and Lorene

PREFACE

This book is an update and revision of my earlier textbook of the same title. The most important change is an increase in the number of worked examples and solved exercises. Also, several new topics have been introduced. But the overall plan of the book is the same as in the first edition: to introduce the reader to the basic elements of combinatorics, along with many examples and exercises.

Combinatorics may be described as the study of how discrete structures can be counted, arranged, and constructed. Accordingly, this book is an introduction to the three main branches of combinatorics: enumeration, existence, and construction. There are two chapters devoted to each of these three areas.

Combinatorics plays a central role in mathematics. One has only to look at the numerous journal titles in combinatorics and discrete mathematics to see that this area is huge! Some of the journal titles are *Journal of Combinatorial Theory Series A* and *Series B; Journal of Graph Theory; Discrete Mathematics; Discrete Applied Mathematics; Annals of Discrete Mathematics; Annals of Combinatorics; Topics in Discrete Mathematics; SIAM Journal on Discrete Mathematics; Graphs and Combinatorics; Combinatorica; Ars Combinatoria; European Journal of Combinatorics A* and *B; Journal of Algebraic Combinatorics; Journal of Combinatorial Designs; Designs, Codes, and Cryptography; Journal of Combinatorial Mathematics and Combinatorial Computing; Combinatorics, Probability & Computing; Journal of Combinatorics, Information & System Sciences; Algorithms and Combinatorics; Random Structures & Algorithms; Bulletin of the Institute of Combinatorics and Its Applications; Journal of Integer Sequences; Geombinatorics; Online Journal of Analytic Combinatorics;* and *The Electronic Journal of Combinatorics*. These journal titles indicate the connections between discrete mathematics and computing, information theory and codes, and probability. Indeed, it is now desirable for all mathematicians, statisticians, and computer scientists to be acquainted with the basic principles of discrete mathematics.

The format of this book is designed to gradually and systematically introduce the main concepts of combinatorics. In this way, the reader is brought step-by-step from first principles to major accomplishments, always pausing to note mathematical points of interest along the way. I have made it a point to discuss some topics that don't receive much treatment in other books on combinatorics, such as Alcuin's sequence, Rook walks, and Leech's lattice. In order to illustrate the applicability of combinatorial methods, I have paid careful attention to the selection of exercises at the end of each section. The reader should definitely attempt the exercises, as a good deal of the subject is revealed there. The problems range in difficulty from very easy to very challenging. Solutions to selected exercises are provided in the back of the book.

I wish to thank the people who have kindly made suggestions concerning this book: Mansur Boase, Robert Cacioppo, Duane DeTemple, Shalom Eliahou, Robert Dobrow, Suren Fernando, Joe Hemmeter, Daniel Jordan, Elizabeth Oliver, Ken Price, Adrienne Stanley, and Khang Tran.

I also gratefully acknowledge the Wiley staff for their assistance in publishing this book: Liz Belmont, Kellsee Chu, Sari Friedman, Danielle LaCourciere, Jacqueline Palmieri, Susanne Steitz-Filler, and Stephen Quigley.

CHAPTER 1

BASIC COUNTING METHODS

We begin our tour of combinatorics by investigating elementary methods for counting finite sets. How many ways are there to choose a subset of a set? How many permutations of a set are there? We will explore these and other such questions.

1.1 The multiplication principle

We start with the simplest counting problems. Many of these problems are concerned with the number of ways in which certain choices can occur.

Here is a useful counting principle: If one choice can be made in x ways and another choice in y ways, and the two choices are independent, then the two choices together can be made in xy ways. This rule is called the "multiplication principle."

EXAMPLE 1.1

Suppose that you have three hats and four scarves. How many different hat and scarf outfits can you choose?

Solution: By the multiplication principle, there are $3 \cdot 4 = 12$ different outfits. Let's call the hats h_1, h_2, and h_3 and the scarves s_1, s_2, s_3, and s_4. Then we can list the different outfits as follows:

$$\begin{array}{cccc} h_1, s_1 & h_1, s_2 & h_1, s_3 & h_1, s_4 \\ h_2, s_1 & h_2, s_2 & h_2, s_3 & h_2, s_4 \\ h_3, s_1 & h_3, s_2 & h_3, s_3 & h_3, s_4 \end{array}$$

EXAMPLE 1.2

At the French restaurant Chacun à Son Goût, there are three choices for the appetizer, four choices for the entrée, and five choices for the dessert. How many different dinner orders (consisting of appetizer, entrée, and dessert) can we make?

Solution: The answer is $3 \cdot 4 \cdot 5 = 60$, and it isn't difficult to list all the possibilities. Let's call the appetizers a_1, a_2, and a_3, the entrées e_1, e_2, e_3, and e_4, and the desserts d_1, d_2, d_3, d_4, and d_5. Then the different possible dinners are as follows:

a_1, e_1, d_1 a_1, e_1, d_2 a_1, e_1, d_3 a_1, e_1, d_4 a_1, e_1, d_5
a_1, e_2, d_1 a_1, e_2, d_2 a_1, e_2, d_3 a_1, e_2, d_4 a_1, e_2, d_5
a_1, e_3, d_1 a_1, e_3, d_2 a_1, e_3, d_3 a_1, e_3, d_4 a_1, e_3, d_5
a_1, e_4, d_1 a_1, e_4, d_2 a_1, e_4, d_3 a_1, e_4, d_4 a_1, e_4, d_5
a_2, e_1, d_1 a_2, e_1, d_2 a_2, e_1, d_3 a_2, e_1, d_4 a_2, e_1, d_5
a_2, e_2, d_1 a_2, e_2, d_2 a_2, e_2, d_3 a_2, e_2, d_4 a_2, e_2, d_5
a_2, e_3, d_1 a_2, e_3, d_2 a_2, e_3, d_3 a_2, e_3, d_4 a_2, e_3, d_5
a_2, e_4, d_1 a_2, e_4, d_2 a_2, e_4, d_3 a_2, e_4, d_4 a_2, e_4, d_5
a_3, e_1, d_1 a_3, e_1, d_2 a_3, e_1, d_3 a_3, e_1, d_4 a_3, e_1, d_5
a_3, e_2, d_1 a_3, e_2, d_2 a_3, e_2, d_3 a_3, e_2, d_4 a_3, e_2, d_5
a_3, e_3, d_1 a_3, e_3, d_2 a_3, e_3, d_3 a_3, e_3, d_4 a_3, e_3, d_5
a_3, e_4, d_1 a_3, e_4, d_2 a_3, e_4, d_3 a_3, e_4, d_4 a_3, e_4, d_5

EXAMPLE 1.3

A variable name in a certain computer programming language consists of a letter (A through Z), a letter followed by another letter, or a letter followed by a digit (0 through 9). How many different variable names are possible?

Solution: There are 26 variable names consisting of a single letter, 26^2 variable names consisting of two letters, and $26 \cdot 10$ variable names consisting of a letter followed by a digit. Altogether, there are

$$26 + 26^2 + 26 \cdot 10 = 962$$

variable names.

EXAMPLE 1.4 Number of binary strings

How many binary strings of length n are there?

Solution: There are two choices (0 or 1) for each element in the string. Hence, there are 2^n possible strings.

For instance, there are $2^3 = 8$ binary strings of lenght 3:

000, 001, 010, 011, 100, 101, 110, 111.

EXAMPLE 1.5 Number of subsets of a set

Let S be a set of n elements. How many subsets does S have?

Solution: There are two choices for each element of S; it can be in the subset or not in the subset. This means that there are 2^n subsets altogether.

For instance, let $S = \{a, b, c\}$, so that $n = 3$. Then S has $2^3 = 8$ subsets:

$\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{b, c\}$, $\{a, c\}$, $\{a, b, c\}$.

EXERCISES

1.1 A person making a book display wants to showcase a novel, a history book, and a travel guide. There are four choices for the novel, two choices for the history book, and 10 choices for the travel guide. How many choices are possible for the three books?

1.2 A license consists of three digits (0 through 9), followed by a letter (*A* through *Z*), followed

by another digit. How many different licenses are there?

1.3 How many strings of length 10 are there in which the symbols may be 0, 1, or 2?

1.4 How many subsets of the set $\{a, b, c, d, e, f, g, h, i, j\}$ do not contain both a and b?

1.5 How many binary strings of length 99 have an odd number of 1′s?

1.6 How many functions map the set $\{a, b, c\}$ to the set $\{w, x, y, z\}$?

1.7 Let X be an n-element set. How many functions from X to X are there?

1.8 Let $X = \{1, 2, 3, \ldots, 2n\}$. How many functions from X to X are there such that each even number is mapped to an even number and each odd number is mapped to an odd number?

1.2 Permutations

One of the fundamental concepts of counting is that of a permutation. A *permutation* of a set is an ordering of the elements of the set.

EXAMPLE 1.6

List the permutations of the set $\{a, b, c\}$.

Solution: There are six permutations:

$$abc,\ acb,\ bac,\ bca,\ cab,\ cba.$$

We set

$$n! = 1 \cdot 2 \cdot 3 \cdots n, \quad n \geq 1; \quad 0! = 1. \tag{1.1}$$

The expression $n!$ is called *n factorial*.

We see in the above example that the number of permutations is $6 = 3!$.

There are $n!$ permutations of an n-element set. The reason is there are n choices for the first element in the permutation. Once that choice is made, there are $n - 1$ choices for the second element, and then $n - 2$ choices for the third element, and so on. Altogether, there are

$$n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1$$

choices, which is $n!$

EXAMPLE 1.7

In how many ways can the letters of the word MISSISSIPPI be arranged?

Solution: This is an example of a permutation of a set with repeated elements. There are 11! permutations of the 11 letters of MISSISSIPPI, but there is much duplication.

We need to divide by the number of permutations of the four I's, the four S's, the two P's, and the one M. Thus, the number of different arrangements of the letters is

$$\frac{11!}{4!4!2!1!} = 34{,}650.$$

Let S be an n-element set, where $n \geq 0$. How many permutations of k elements of S are there, where $1 \leq k \leq n$? There are n choices for the first element, $n - 1$ choices for the second element,…,$n - k + 1$ choices for the kth element. Hence, there are

$$n(n-1)\cdots(n-k+1)$$

choices altogether. This expression, denoted $P(n, k)$, may be written as

$$P(n,k) = \frac{n!}{(n-k)!}, \quad 0 \le k \le n. \tag{1.2}$$

(Notice that we now allow $k = 0$, which gives $P(n, 0) = 1$.) We can interpret this formula as a MISSISSIPPI-type problem. The selected elements may be denoted $X_1,\ldots,X_k$, and the nonselected elements all denoted with the letter N (for nonselected).

EXAMPLE 1.8

An organization has 100 members. How many ways may they select a president, a vice-president, a secretary, and a treasurer?

Solution: The number of ways to select a permutation of four people from a group of 100 is

$$P(100, 4) = 100 \cdot 99 \cdot 98 \cdot 97 = 94{,}109{,}400.$$

EXERCISES

1.9 A teacher has eight books to put on a shelf. How many different orderings of the books are possible?

1.10 You have three small glasses, four medium-size glasses, and five large glasses. If glasses of the same size are indistinguishable, how many ways can you arrange the glasses in a row?

1.11 A couple plans to visit three selected cities in Germany, followed by four selected cities in France, followed by five selected cities in Spain. In how many ways can the couple order their itinerary?

1.12 A student has 10 books but only room for six of them on a shelf. How many permutations of the books are possible on the shelf?

1.13 A librarian wants to arrange four astronomy books, five medical books, and six religious books on a shelf. Books of the same category should be grouped together, but otherwise the books may be put in any order. How Many orderings are possible?

1.14 In how many ways can you arrange the letters of the word RHODODENDRON?

1.15 How many one-to-one functions are there from the set $\{a, b, c\}$ to the set $\{t, u, v, w, x, y, z\}$?

1.16 Let X be an n-element set. How many functions from X to X are not one-to-one?

1.17 Find a formula for the number of different binary relations possible on a set of n elements.

1.3 Combinations

Another fundamental concept of counting is that of a combination. A *combination* from a set is an unordered subset (of a given size) of the set.

For convenience, we sometimes refer to an n-element set as an n-set and a k-element subset as a k-subset. Also, we use the notation $\mathbf{N} = \{1, 2, 3,\ldots,\}$ and $\mathrm{N}_m = \{1, 2, 3,\ldots,m\}$.

Let S be an n-set, where $n \ge 0$. How many k-subsets of S are there, where $0 \le k \le n$? We can regard this as a MISSISSIPPI-type problem, i.e., a problem of permutations with repeated elements. Let X denote selected elements and N denote nonselected elements. Then the number of combinations is the

number of arrangements of k X's and $n - k$ N's, since each such arrangement specifies a combination.

Hence, the number of combinations, denoted $C(n, k)$, is given by

$$C(n,k) = \frac{n!}{k!(n-k)!}, \quad 0 \le k \le n. \tag{1.3}$$

We call this expression "n choose k." We set $C(n, k) = 0$ for $k < 0$ and $k > n$.

For example, with $n = 5$ and $k = 3$, we have $C(5, 3) = 5!/(3!2!) = 10$ combinations of three elements from the set $S = \{a, b, c, d, e\}$, as shown below with the corresponding arrangements of X's and N's:

$\{a,b,c\}$	$\{a,b,d\}$	$\{a,b,e\}$	$\{a,c,d\}$	$\{a,c,e\}$
XXXNN	XXNXN	XXNNX	XNXXN	XNXNX
$\{a,d,e\}$	$\{b,c,d\}$	$\{b,c,e\}$	$\{b,d,e\}$	$\{c,d,e\}$
XNNXX	NXXXN	NXXNX	NXNXX	NNXXX.

The values of $C(n, k)$ are given by a famous array of numbers known as *Pascal's triangle*. See Figure 1.1. The triangle is created by starting with a 1 in the top row, placing 1′s at the ends of each successive row, and adding two consecutive entries in a row to produce the entry beneath and between these entries. Thus, we can generate Pascal's triangle from the initial values

Figure 1.1 Pascal's triangle.

```
                                1
                              1   1
                            1   2   1
                          1   3   3   1
                        1   4   6   4   1
                      1   5   10  10  5   1
                    1   6   15  20  15  6   1
                  1   7   21  35  35  21  7   1
                1   8   28  56  70  56  28  8   1
              1   9   36  84  126 126 84  36  9   1
            1   10  45  120 210 252 210 120 45  10  1
                                ⋮
```

$$C(n,0) = 1 \quad \text{and} \quad C(n,n) = 1 \quad \text{for all } n \ge 0 \tag{1.4}$$

and the relation

$$C(n,k) = C(n-1,k-1) + C(n-1,k), \quad 1 \le k \le n-1. \tag{1.5}$$

The rows of Pascal's triangle arc numbered 0, 1, 2, etc. (from top to bottom), and the columns are numbered 0, 1, 2, etc. (from left to right). The entry in row n, column k of Pascal's triangle is $C(n, k)$.

EXAMPLE 1.9

Evaluate $C(10, 5)$.

Solution: We see that the 5th entry of the 10th row of Pascal's triangle is 252. Hence $C(10, 5) = 252$. This means that there are 252 combinations of five objects from a set of 10 objects.

Pascal's triangle gives the coefficients of the expansion of a binomial, such as $a + b$, raised to a power. For example,

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

and we see that the coefficients 1, 3, 3, 1 constitute the third row of Pascal's triangle. For this reason, the entries of Pascal's triangle are called *binomial coefficients*.

We often use the notation $\binom{n}{k}$ for $C(n, k)$. We may also write this binomial coefficient as $\binom{n}{k,n-k}$. Thus, we know

$$\binom{n}{0} = 1 \quad \text{and} \quad \binom{n}{n} = 1 \quad \text{for all } n \geq 0 \tag{1.6}$$

and

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad 1 \leq k \leq n-1. \tag{1.7}$$

We also have the formula

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad 0 \leq k \leq n. \tag{1.8}$$

Binomial theorem. For any numbers a and b (real or complex) and any nonnegative integer n, we have

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

Proof. We give a combinatorial proof showing that, for each k, where $0 \leq k \leq n$, the coefficients of $a^{n-k}b^k$ on the two sides of the equation are equal. The coefficient $a^{n-k}b^k$ on the left side is the number of ways of selecting $n - k$ factors $a + b$ that contribute a's to the expansion of $(a + b)^n$ (the other factors contribute b's). Such selections are choices of k unordered objects from a set of n objects; hence there are $\binom{23}{7,10,6}$ of them. This number is the coefficient of $a^{n-k}b^k$ on the right side.

We could also have proved the binomial theorem by mathematical induction using (1.6) and (1.7).

EXAMPLE 1.10

What is the coefficient of $a^{12}b^8$ in the expansion of $(a + b)^{20}$?

Solution: By the binomial theorem, the coefficient is

$$\binom{20}{8} = \frac{20!}{8!12!} = 125{,}970.$$

How can we find the expansion of a multinomial expression raised to a power, such as $(a + b + c)^{10}$? The answer is given by the multinomial theorem.

From the solution to the MISSISSIPPI problem, we know that the number of ways that n objects can be divided into groups of sizes $k_1, k_2, \ldots, k_m$, such that $k_1 + k_2 + \cdots + k_m = n$, where order among and within groups is unimportant, is

$$\frac{n!}{k_1!k_2!\cdots k_m!}.$$

This expression, called a *multinomial coefficient*, is denoted by

$\binom{n}{k_1, k_2, \ldots, k_m}$.

Multinomial theorem. In the expansion of $(x_1 + x_2 + \cdots + x_m)^n$, the coefficient of $(x_1 + x_2 + \cdots + x_m)^n$, where the k_i are nonnegative integers such that $k_1 + k_2 + \cdots + k_m = n$, is the multinomial coefficient

$$\binom{n}{k_1, k_2, \ldots, k_m}.$$

EXAMPLE 1.11

Give the expansion of $(a + b + c)^3$.

Solution: By the multinomial theorem,

$$\begin{aligned}(a+b+c)^3 &= \binom{3}{3,0,0}a^3 + \binom{3}{2,1,0}a^2b + \binom{3}{1,2,0}ab^2 \\ &\quad + \binom{3}{0,3,0}b^3 + \binom{3}{2,0,1}a^2c + \binom{3}{1,1,1}abc \\ &\quad + \binom{3}{0,2,1}b^2c + \binom{3}{1,0,2}ac^2 + \binom{3}{0,1,2}bc^2 + \binom{3}{0,0,3}c^3 \\ &= a^3 + 3a^2b + 3ab^2 + b^3 + 3a^2c + 6abc + 3b^2c + 3ac^2 + 3bc^2 + c^3.\end{aligned}$$

We can also think of a multinomial coefficient as an "ordered partition." For instance, the multinomial coefficient $\binom{23}{7,10,6}$ is the number of partitions of the set $\{1, 2, 3, \ldots, 23\}$ into three subsets, *A, B*, and *C*, where *A* has seven elements, *B* has ten elements, and *C* has six elements.

EXERCISES

1.18 A student decides to take three classes from a set of 10. In how many ways may she do this?

1.19 Evaluate $C(20, 10)$.

1.20 Give the expansion of $(a + b)^{10}$.

1.21 What is the coefficient of $a^{10}b^{10}$ in the expansion of $(a + b)^{20}$?

1.22 Give simple formulas for $\binom{n}{1}$, $\binom{n}{2}$ and $\binom{n}{3}$.

1.23 Explain, in terms of counting, the formula

$$C(n, k) = \frac{P(n, k)}{k!}.$$

1.24 A pointer starts at 0 on the real number line and moves right or left one unit at each step. Let *n* and *k* be positive integers. How many different paths of *k* steps terminate at the integer *n*?

1.25 Give the expansion of $(a + b + c)^4$.

1.26 What is the coefficient of x^3y^7 in the expansion of $(x + y + 1)^{20}$?

1.27 Show that the multinomial coefficient

$$\binom{n}{k_1, k_2, \ldots, k_m}$$

is equal to a product of binomial coefficients.

1.28 Prove the following relations for multinomial coefficients:

$$\binom{n}{k_1,k_2,k_3} = \binom{n}{k_1-1,k_2,k_3} + \binom{n}{k_1,k_2-1,k_3} + \binom{n}{k_1,k_2,k_3-1},$$
$$k_1,k_2,k_3 \geq 1$$

$$\binom{n}{0,k_2,k_3} = \binom{n}{k_2,k_3}$$

$$\binom{n}{k_1,0,k_3} = \binom{n}{k_1,k_3}$$

$$\binom{n}{k_1,k_2,0} = \binom{n}{k_1,k_2}.$$

1.29 Prove the multinomial theorem.

1.30 (a) How many paths in $\mathbf{R}^2$ start at the origin (0, 0), move in steps of (1, 0) or (0, 1), and end at (10, 15)?

(b) How many paths in $\mathbf{R}^3$ start at the origin (0, 0, 0), move in steps of (1, 0, 0), (0, 1, 0), or (0, 0, 1), and end at (10, 15, 20)?

1.4 Binomial coefficient identities

Looking at Pascal's triangle (Figure 1.1), we see quite a few patterns. Notice that the triangle is symmetric about a vertical line down the middle. To prove this, let X be an n-set. Then a natural bijection between the collection of k-subsets of X and the collection of $(n-k)$-subsets of X (simply pair each subset with its complement) shows that the two binomial coefficients in question are equal:

$$\binom{n}{k} = \binom{n}{n-k}. \tag{1.9}$$

This identity also follows instantly from the formulas for $\binom{n}{k}$ and $\binom{n}{n-k}$.

Many identities can be proved both algebraically and combinatorially. Often, the combinatorial proof is more transparent.

The rule that generates Pascal's triangle (together with the values $\binom{n}{0} = \binom{n}{n} = 1$) is known as *Pascal's identity.*

Pascal's identity.

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad 1 \leq k \leq n-1.$$

Pascal's identity has a simple combinatorial proof. The binomial coefficient $\binom{n}{k}$ is the number of k-subsets of the set $\{1,\ldots,n\}$. Each such subset either contains the element 1 or does not contain 1. The number of k-subsets that contain 1 is $\binom{n-1}{k-1}$. The number of k-subsets that do not contain 1 is $\binom{n-1}{k}$. The identity follows from this observation.

The combinatorial proof of Pascal's identity is more enlightening than the following algebraic derivation:

$$\begin{aligned}
\binom{n-1}{k-1}+\binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!(n-k)!}+\frac{(n-1)!}{k!(n-k-1)!} \\
&= \frac{(n-1)!\cdot k}{k!(n-k)!}+\frac{(n-1)!\cdot(n-k)}{k!(n-k)!} \\
&= \frac{(n-1)!\cdot(k+n-k)}{k!(n-k)!} \\
&= \frac{(n-1)!\cdot n}{k!(n-k)!} \\
&= \frac{n!}{k!(n-k)!} \\
&= \binom{n}{k}.
\end{aligned}$$

EXAMPLE 1.12 Sum of a row of Pascal's triangle

The sum of the entries of the nth row of Pascal's triangle is 2^n.

$$\sum_{k=0}^{n}\binom{n}{k}=2^n \tag{1.10}$$

Combinatorially, this identity says that the number of subsets of an n-set is equal to the number of k-subsets of the n-set, summed over all $k = 0,\ldots n$. The identity also follows by putting $a = b = 1$ in the binomial theorem.

EXAMPLE 1.13 Alternating sum of a row of Pascal's triangle

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

Solution: We give three solutions. (1) Letting $a = -1$ and $b = 1$ in the binomial theorem, we obtain

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}=0^n=\begin{cases}1 \text{ for } n=0\\ 0 \text{ for } n>0.\end{cases} \tag{1.11}$$

(2) Here is a combinatorial proof. Consider the equivalent formulation

$$\sum_{k \text{ odd}}\binom{n}{k}=\sum_{k \text{ even}}\binom{n}{k}.$$

This relation says that, for any $n > 0$, the number of subsets of $X = \{1,\ldots,n\}$ with an odd number of elements is equal to the number of subsets with an even number of elements. For n odd, this assertion follows trivially from the symmetry of the binomial coefficients. We give a combinatorial argument valid for any $n > 0$. Let

$$\begin{aligned}
\mathcal{A} &= \{S\subseteq X : |S| \text{ is even and } 1\in S\} \\
\mathcal{B} &= \{S\subseteq X : |S| \text{ is odd and } 1\in S\} \\
\mathcal{C} &= \{S\subseteq X : |S| \text{ is even and } 1\notin S\} \\
\mathcal{D} &= \{S\subseteq X : |S| \text{ is odd and } 1\notin S\}.
\end{aligned}$$

The obvious bijections between $\mathcal{A}$ and $\mathcal{D}$ and between $\mathcal{B}$ and $\mathcal{C}$ establish that $|\mathcal{A}| = |\mathcal{D}|$ and $|\mathcal{B}| = |\mathcal{C}|$, and hence

$|\mathcal{A}| + |\mathcal{C}| = |\mathcal{B}| + |\mathcal{D}|$.

The identity follows immediately.

(3) The identity can be turned into a telescoping series. For $n > 0$, we have

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k} = \sum_{k=0}^{n}(-1)^k\left[\binom{n-1}{k-1} + \binom{n-1}{k}\right] = 0.$$

EXAMPLE 1.14 Sum of squares of a row of Pascal's triangle

What is the sum $\sum_{k=0}^{n}\binom{n}{k}^2$?

Solution: Let's work out some instances of the sum using Pascal's triangle:

$$n = 1:\ 1^2 = 1$$
$$n = 2:\ 1^2 + 1^2 = 2$$
$$n = 3:\ 1^2 + 2^2 + 1^2 = 6$$
$$n = 4:\ 1^2 + 3^2 + 3^2 + 1^2 = 20$$
$$n = 5:\ 1^2 + 4^2 + 6^2 + 4^2 + 1^2 = 70.$$

We recognize these sums as central binomial coefficients and conjecture that

$$\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}. \tag{1.12}$$

Typically, the mathematical process consists of working example, looking for patterns, making conjectures, and proving the conjectures. Let's try to prove our conjecture.

We rewrite our conjecture as follows:

$$\binom{n}{0}\binom{n}{0} + \binom{n}{1}\binom{n}{1} + \binom{n}{2}\binom{n}{2} + \cdots + \binom{n}{n}\binom{n}{n} = \binom{2n}{n}.$$

We know that the right side counts the ways of selecting n numbers from the set $\{1,2,3,\ldots,2n\}$. Why is this counted by the left side? Rewrite just a little, using Symmetry:

$$\binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \binom{n}{2}\binom{n}{n-2} + \cdots + \binom{n}{n}\binom{n}{0} = \binom{2n}{n}.$$

Now the truth of the identity is clear. The right side counts the number of n-subsets of $\{1,2,3,\ldots,2n\}$. The left side counts the same thing, according to the number of elements that are chosen from the subset $\{1,2,3,\ldots,n\}$.

This identity has an interesting combinatorial interpretation. The binomial coefficient $\binom{2n}{n}$ is the number of northeast paths which start at the southwest corner of an $n \times n$ grid and stop at the northeast corner. Such paths are of length $2n$ and are determined by a sequence of n "easts" and n "norths" in some order. The summation $\sum_{i=0}^{n}\binom{n}{i}^2$ counts the paths according to their intersection with the main diagonal of the grid. The number of paths that cross the diagonal at the points i units east of the starting point is $\binom{n}{i}^2$, where $0 \le i \le n$.

Other binomial coefficient identities may be obtained by comparing like powers of x in certain algebraic identities. For example, comparing coefficients of x^k in the polynomial identity

$$(x+1)^{m+n} = (x+1)^m(x+1)^n,$$

we obtain *Vandermonde's identity*.

Vandermonde's identity.

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}.$$

Vandermonde's identity has a combinatorial interpretation. The binomial coefficient $\binom{m+n}{k}$ is the number of k-subsets of the $(m + n)$-set $A \cup B$, where $A = \{1,\ldots,m\}$ and $B = \{m + 1,\ldots,m + n\}$. The number of such subsets that contain i elements of A (and $k - i$ elements of B) is $\binom{m}{i}\binom{n}{k-i}$, and the summation $\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}$ counts these subsets for $i = 1,\ldots,k$.

Letting $m = 1$, and changing n to $n - 1$, the relation becomes Pascal's identity. Putting $k = m = n$, we obtain a previously seen identity:

$$\binom{2n}{n} = \sum_{i=0}^{n} \binom{n}{i}^2.$$

Here is another algebraic identity:

$$(x+1)^{m+n+1} = \underbrace{(x+1)\cdots(x+1)}_{m+n+1}.$$

The coefficient of x^{n+1} on the left side is $\binom{m+n+1}{n+1}$. On the right side, there is a contribution to x^{n+1} whenever we multiply x's from $n + 1$ of the factors. Suppose that the rightmost factor which contributes an x is the $(n + i + 1)$st factor, where $0 \le i \le m$. This leaves us free to choose n other x's from a set of $n + i$ factors. Hence the coefficient of x^{n+1} on the right side is $\sum_{i=0}^{m} \binom{n+i}{n}$. This proves the identity

$$\binom{m+n+1}{n+1} = \sum_{i=0}^{m} \binom{n+i}{n}. \tag{1.13}$$

The identity has a combinatorial interpretation. The binomial coefficient $\binom{m+n+1}{n+1}$ is the number of $(n+1)$-subsets of the $(m+n+1)$-set $\{1,\ldots,m+n+1\}$. Suppose that the largest element in such a subset is $n + i + 1$, where $0 \le i \le m$. The number of such subsets is $\binom{n+i}{n}$, and the summation $\sum_{i=0}^{m} \binom{n+i}{n}$ counts them all.

Subcommittee identity. For $0 \le j \le k \le n$, we have

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

Proof. Both expressions count the number of ways to choose, from n people, a committee of size k and a subcommittee of size j.

EXAMPLE 1.15

Prove the identity:

$$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}.$$

Solution: We will give four proofs.

(1) The first proof is algebraic. We can "pull an n out" of each term in the sum to obtain

$$\begin{aligned}\sum_{k=1}^{n} k\binom{n}{k} &= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} \\ &= n\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} \\ &= n\sum_{k=1}^{n} \binom{n-1}{k-1} \\ &= n2^{n-1}.\end{aligned}$$

(2) The second proof is by counting. Consider all possible ways of choosing a team and a team leader from a set of n people. The left side clearly counts this, according to the size k of the team. The right side counts the same thing, as we have n choices for the leader and each other person can be on or off the team.

(3) The third proof uses calculus. From the binomial theorem, we have

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

Taking a derivative "brings a k down," so

$$\frac{d}{dx}(x+1)^n = n(x+1)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} kx^{k-1}.$$

Evaluating both sides of the last relation at $x = 1$ gives our desired identity.

(4) Let's also do a proof via probability. Upon division by 2^n, our identity becomes

$$\sum_{k=1}^{n} k\binom{n}{k}\left(\frac{1}{2}\right)^n = \frac{n}{2}.$$

Here is a probabilistic interpretation. Let X be a set of n elements. For each element of X, flip a fair coin and if the coin comes up heads put the element in a subset S. What is the expected size of S? Both sides of the identity give the answer!

EXAMPLE 1.16

Prove the identity

$$\sum_{k=0}^{n} \binom{n+k}{n} 2^{-k} = 2^n.$$

Solution: We give a counting proof of the equivalent identity

$$\sum_{k=0}^{n} \binom{n+k}{n} 2 \cdot 2^{n-k} = 2^{2n+1}.$$

The right side of this relation is the number of binary strings of length $2n + 1$. We must show that the left side counts the same strings. Every binary string of length $2n + 1$ contains at least $n + 1$ 0′s or at least $n+ 1$ 1′s(but not both). Counting from the left, let $n+ k+1$, where $0 \le k \le < n$, be the position of the $(n +1)$st 0 or $(n +1)$st 1. There are two possibilities for this element (0 or 1); there are $\binom{n+k}{n}$ binary strings of length $n + k$ that contain n of one symbol and k of the other; and there are 2^{n-k} choices for the remaining $n - k$ elements. This establishes the identity.

EXAMPLE 1.17 An object moving in the plane

An object travels along the integer points of the plane, starting at the point (0, 0). At each step, the object moves one unit to the right or one unit up (with equal probability). The object stops when it reaches the line $x = n$ or the line $y = n$. Show that the expected length of the object's path is $2n - n\binom{2n}{n}2^{1-2n}$.

Solution: Assume that the object hits the line $x = n$ at the point (n, k) or the line $y = n$ at the point (k, n), where $0 \le k \le n - 1$. Then the expected path length is given by

$$\begin{aligned} E &= \sum_{k=0}^{n-1}(n+k)2 \cdot \frac{1}{2}\binom{n+k-1}{n-1}\left(\frac{1}{2}\right)^{n+k-1} \\ &= \left(\frac{1}{2}\right)^{n}\sum_{k=0}^{n-1}(n+k)\binom{n+k-1}{n-1}\left(\frac{1}{2}\right)^{k-1} \\ &= \left(\frac{1}{2}\right)^{n} n\sum_{k=0}^{n-1}\binom{n+k}{n}\left(\frac{1}{2}\right)^{k-1} \\ &= \left(\frac{1}{2}\right)^{n} 2n\sum_{k=0}^{n-1}\binom{n+k}{n}\left(\frac{1}{2}\right)^{k}. \end{aligned}$$

By the result of Example 1.16, this simplifies to

$$\begin{aligned} E &= \left(\frac{1}{2}\right)^{n} 2n\left[\sum_{k=0}^{n}\binom{n+k}{n}\left(\frac{1}{2}\right)^{k} - \binom{2n}{n}\left(\frac{1}{2}\right)^{n}\right] \\ &= 2n - n\binom{2n}{n}2^{1-2n}. \end{aligned}$$

The binomial theorem extends to arbitrary exponents. For any real number α and k a positive integer, define

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}. \tag{1.14}$$

Also, define $\binom{\alpha}{0} = 1$.

EXAMPLE 1.18

$$\binom{-3}{4} = \frac{(-3)(-4)(-5)(-6)}{4!} = 15$$

Binomial series. Let α be a real number and $|x| < 1$. Then

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty}\binom{\alpha}{k}x^{k}.$$

EXAMPLE 1.19

Let n be an integer greater than or equal to 1. Prove the formula

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

Solution: By the binomial series theorem,

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

The result now follows from the identity (see Exercises)

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}. \tag{1.15}$$

Pascal's triangle extends upward, as in Figure 1.2. (In the figure, the triangle is left-justified.) Pascal's identity, together with the initial values $\binom{n}{0} = 1$ for all integers n, is used to calculate entries $\binom{-n}{k}$, where n is a positive integer. The entries are the numbers $(-1)^k \binom{n+k-1}{k}$ given by the binomial series theorem as the coefficients of x in the power series expansion of $(1+x)^{-n}$.

Figure 1.2 The extended Pascal's triangle.

⋮	⋮	⋮	⋮	⋮	⋮	
1	−5	15	−35	70	−126	⋯
1	−4	10	−20	35	−56	⋯
1	−3	6	−10	15	−21	⋯
1	−2	3	−4	5	−6	⋯
1	−1	1	−1	1	−1	⋯
1	0	0	0	0	0	⋯
1	1	0	0	0	0	⋯
1	2	1	0	0	0	⋯
1	3	3	1	0	0	⋯
1	4	6	4	1	0	⋯
1	5	10	10	5	1	⋯
⋮	⋮	⋮	⋮	⋮	⋮	

EXAMPLE 1.20

Give the first several terms of the expansion of $(1+x)^{-4}$ in powers of x.

Solution: We can see the coefficients in row −4 of the extended Pascal's triangle. Thus

$$(1+x)^{-4} = 1 - 4x + 10x^2 - 20x^3 + 35x^4 - 56x^5 + \cdots.$$

EXERCISES

1.31 Prove the following identities:

(a) $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$

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

1.32 Prove the identity

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n \cdot (n+1) = \frac{n(n+1)(n+2)}{3}.$$

Generalize.

1.33 Prove the following identities:

(a) $\binom{2n-1}{n} = \sum_{k=0}^{n} \binom{n}{k}\binom{n-1}{k}$

(b) $\binom{3n}{n} = \sum_{k=0}^{n} \binom{n}{k}\binom{2n}{k}$

1.34 (a) Prove the identity $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$.

(b) Use the identity of part (a) to show that the entries of each row of Pascal's triangle increase from left to right, attain a maximum value at the middle entry (or two middle entries), and then decrease.

1.35 Prove the inequality $\binom{n}{k}^2 \geq \binom{n}{k-1}\binom{n}{k+1}$, where $1 \leq k \leq n-1$.

1.36 Suppose that five particles are traveling back and forth on the unit interval [0, 1]. Initially, all the particles move to the right with the same speed. (The initial placement of the particles does not matter, as long as they are not at the endpoints.) When a particle reaches 0 or 1, it reverses direction but maintains its speed. When two particles collide, they both reverse direction (and maintain speeds). How many particle—particle collisions occur before the particles once again occupy their original positions and are moving to the right?

1.37 Show that $2n$ people may be grouped into n pairs in $(2n)!/(n!2^n)$ ways.

1.38 How many ways can $3n$ people be grouped into n trios?

1.39 How many ways can kn people be grouped into it subgroups of size k?

1.40 Prove that the number of binary strings of length n that contain exactly k copies of the string 10 is

$$\binom{n+1}{2k+1}.$$

1.41 Give the first several terms of the expansion of $(1+x)^{1/2}$ in powers of x.

1.42 Give the first several terms of the expansion of $(1+x)^{-5}$ in powers of x.

1.43 Give the first several terms of the expansion of $(1+x)^{-1/2}$ in powers of x.

1.44 Prove the identity

$$\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{n+j+k}{n,j,k}\left(\frac{1}{3}\right)^{j+k} = 3^n.$$

1.45 For each integer $k \geq 0$, define

$$S_k(n) = \sum_{i=1}^{n} i^k.$$

Give formulas for $S_0(n)$, $S_1(n)$, $S_2(n)$, and $S_3(n)$. Prove that $S_k(n)$ is a polynomial in n of degree $k + 1$ and leading coefficient $1/(k+1)$.

1.46 An n-dimensional *hypercube* consists of all binary n-tuples. Two such n-tuples are joined by an edge if they disagree in exactly one coordinate Prove that the number of k-dimensional faces of an n-dimensional hypercube is

$$\binom{n}{k}2^{n-k}, \quad 0 \leq k \leq n.$$

1.5 Distributions

Problems in which elements of a set are divided into categories are called *distribution problems*. Let's consider a simple scenario. Suppose that five $1 bills are to be distributed among three people. In how many ways can this be done? The answer depends on whether the people are to be considered as identifiable in some way, and the same goes for the dollar bills. For instance, suppose that the people are named Amy, Bobby, and Carly, and the dollar bills have serial numbers so they are identifiable. Then there are three choices for who gets the first dollar bill, three choices for who gets the second dollar bill, and so on. Altogether, there are 3^5 ways to distribute the five dollars to the three people.

If the dollar bills are interchangeable, then we have a so-called "stars and bars" situation. The number of distributions is the number of ways to arrange five dollar signs (or stars) and two vertical lines (or bars) partitioning the three people along a line. The number of ways is $C(7, 2)$.

If the three people are anonymous but the five bills are numbered, then the number of distributions is given by the Stirling numbers of the second kind. We will see more about this later in this section.

If the people are anonymous and the bills are interchangeable, then we have what are called partition numbers. We will sec mom about this later, too.

EXAMPLE 1.21

How many solutions in nonnegative integers are there to the equation

$$x_1 + x_2 + x_3 = 10?$$

Solution: We can think of the 10 on the right side of the equation as representing 10 units that can be distributed to the three variables, x_1, x_2, and x_3. Such a distribution can be pictured with a linear ordering of 10 *'s (to represent the units) and two vertical lines (to indicate the partitioning of the units among the variables). For instance, the solution 3 + 2 + 5 = 10 is shown as

$$***|**|*****.$$

Thus, finding the number of distributions is a MISSISSIPPI-type problem. As there are 12 symbols altogether (10 *'s and two vertical lines), the number of solutions is

$$\frac{12!}{10!2!} = 66.$$

Distribution of identical objects into distinguishable classes. The number of ways to distribute k identical objects among n distinguishable classes is $\binom{n+k-1}{k}$. This is the same as the number of nonnegative integer solutions to

$$x_1 + x_2 + \cdots + x_n = k.$$

By contrast, the number of ways to distribute k distinguishable objects into n distinguishable classes is n^k.

A *partition* of X is a collection C of nonempty pairwise-disjoint subsets of X whose union equals X. The members of C are called the *parts* of the partition.

An *equivalence relation* on X is a relation on X that is reflexive, symmetric, and transitive. If R is an equivalence relation on X, then, for each $a \in X$, the set $[a] = \{b \in, X : (a, b) \in R\}$ is the *equivalence*

class of *a*.

Equivalence of equivalence relations and partitions. Let X be a nonempty set. The equivalence classes of an equivalence relation on X are the parts of a partition of X. Conversely, the parts of a partition of X are the equivalence classes of an equivalence relation on X.

Proof. Given an equivalence relation R on X, we will show that $C = \{[x] : x \in X\}$ is a partition of X. First, each member $[x]$ of C is nonempty (it contains x). Second, the union of the members of C is all of X, since each element $x \in X$ is contained in a member of C, namely, $[x]$. Third, the members of C are disjoint. For suppose that $[x] \cap [y]$ is nonempty for some $x, y \in X$; assume that $z \in [x] \cap (y)$. Then, since $(x, z) \in R$ and $(y, z) \in R$, it follows by symmetry and transitivity that $(y, x) \in R$. Let x' be an arbitrary element of $[x]$. Then, since $(x, x') \in R$, it follows by symmetry and transitivity that $(y, x') \in R$, and hence $x' \in [y]$. Since x' is an arbitrary element of $[x]$, we conclude that $[x] \subseteq [y]$. A similar argument shows that $[y] \subseteq [x]$ and therefore $[x] = [y]$.

Now suppose that $\mathcal{C}$ is a partition of X, and define a relation R on X so that $(x, y) \in R$ if $x, y \in C$ for some $C \in \mathcal{C}$. We will show that R is an equivalence relation on X. Since C is a partition of X, each $x \in X$ is an element of some member of $\mathcal{C}$; hence R is reflexive. If x and y are both elements of some member C of $\mathcal{C}$, then the same can be said of y and x; hence R is symmetric. As for transitivity, if x and y are both elements of C for some $C \in \mathcal{C}$, and y and z are both elements of D for some $D \in \mathcal{C}$, then $C = D$ (since the parts of a partition are disjoint). Hence, x and z are both elements of the same member of $\mathcal{C}$.

EXAMPLE 1.22

How many partitions of the set {1, 2, 3, 4} are there?

Solution: In a partition of a set of four elements, the sizes of the equivalence classes sum to 4. There are five possibilities for these sizes:

$$4$$
$$3+1$$
$$2+2$$
$$2+1+1$$
$$1+1+1+1.$$

For example, the partition

$$\{\{1,2\},\{4\},\{5\}\}$$

is of type 2 + 1 + 1. It is an easy matter to count the partitions of each type, obtaining, respectively, 1, 4, 3, 6, and 1, for a total of 15.

The nth Bell number, denoted $B(n)$, is the total number of partitions of the set (1, 2, 3,…,n). The *Stirling number of the second kind* $\left\{ {n \atop k} \right\}$ is the number of partitions of {1, 2, 3,…,n} into k equivalence classes. The *partition number* $p(n)$ is the total number of partitions of a set of n indistinguishable elements. These are also called partitions of an integer.

According to the above example, $B(4) = 15$, $\left\{ {4 \atop 1} \right\} = 1$, $\left\{ {4 \atop 2} \right\} = 7$, $\left\{ {4 \atop 3} \right\} = 6$, $\left\{ {4 \atop 4} \right\} = 1$, and $p(4) = 5$.

We also define $p(n, k)$ to be the number of partitions of n indistinguishable objects into k parts. From the above example, $p(4, 1) = 1$, $p(4, 2) = 2$, $p(4, 3) = 1$, and $p(4, 4) = 1$.

EXERCISES

1.47 You can order a pizza with up to four toppings (repetitions allowed) from a set of 12 toppings. The order of the toppings is unimportant. How many different pizzas can you order?

1.48 In how many ways may k indistinguishable balls be placed in n distinguishable urns so that each urn contains an odd number of balls?

1.49 (a) Find a formula for the number of functions $f : \mathrm{N}_m \to \mathrm{N}_n$ with the property that $f(x) < f(y)$ whenever $1 \le x < y \le m$.

(b) Find a formula for the number of functions $f : \mathrm{N}_m \to \mathrm{N}_n$ with the property that $f(x) \le f(y)$ whenever $1 \le x < y \le m$.

1.50 Find $\left\{ {5 \atop 1} \right\}$, $\left\{ {5 \atop 2} \right\}$, $\left\{ {5 \atop 3} \right\}$, $\left\{ {5 \atop 4} \right\}$, $\left\{ {5 \atop 5} \right\}$, and $B(5)$.

1.51 Find $p(5, 1)$, $p(5, 2)$, $p(5, 3)$, $p(5, 4)$, and $p(5)$.

1.52 Prove that $\left\{ {n \atop 2} \right\} = 2^{n-1} - 1$ and $\left\{ {n \atop n-1} \right\} = \binom{n}{2}$ for $n \ge 2$.

1.53 Determine the number of nonnegative integer solutions to the equation

$$a + 2b + 4c = 10^{30}.$$

1.54 Let $S(n) = |\{(k_1, \ldots, k_m) : m, k_i \in \mathbf{N},\ \sum_{i=1}^{m} k_i = n\}|$. Find with proof a formula for $S(n)$. Note that $S(n)$ counts the number of ways n may be written as $n = k_1 + \cdots + k_m$ for any m (order important). Such summations are called *compositions* of n.

1.55 How many commutative groups of order one million are there?

1.6 The principle of inclusion and exclusion

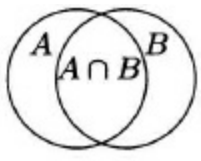
The inclusion–exclusion principle is a generalization of the familiar Venn diagram rule.

Venn diagram rule. If A and B are finite sets, then

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{1.16}$$

Proof. See Figure 1.3, which shows two sets, A and B, and their union and intersection. The sum |A| + |B| counts all the elements of $A \cup B$, but the elements of $A \cap B$ are counted twice and therefore must be removed as on the right side of (1.16).

Figure 1.3 A Venn diagram for two sets.



$A \cup B$

Inclusion–exclusion principle. If $A_1,\ldots,A_n$ are subsets of a finite set S, then

$$|A_1 \cup \cdots \cup A_n| = \sum_{i=1}^{n} (-1)^{i+1} \sum |A_{k_1} \cap \cdots \cap A_{k_i}|, \tag{1.17}$$

where the second sum is over all i-tuples $(k_1,\ldots,k_i)$ with $1 \le k_1 < \cdots < k_i \le n$.

Proof. Let $s \in S$ and assume that s is contained in exactly m of the A_i. The contribution of s to the right side of (1.17) is 0 if $m = 0$. If $m \ge 1$, then the contribution is

$$\begin{aligned}
\sum_{i=1}^{n} (-1)^{i+1} \binom{m}{i} &= \sum_{i=1}^{m} (-1)^{i+1} \binom{m}{i} \qquad \text{(because } m \le n\text{)} \\
&= (-1)\left[\sum_{i=0}^{m} \binom{m}{i} (-1)^i - 1\right] \\
&= (-1)[(-1+1)^m - 1] \\
&= 1.
\end{aligned}$$

Therefore, each $s \in S$ not in the union of the A_i contributes zero to both sides of (1.17), while each $s \in S$ in the union contributes 1. This means that each element of S contributes an equal amount to both sides of (1.17); hence, (1.17) is a valid relation.

EXAMPLE 1.23 Derangements

A permutation with no fixed points is called a *derangement*. Let d_n be the number of derangements of n elements. Find a formula for d_n.

Solution: For $1 \le j \le n$, let A_j be the set of permutations of $\{1, 2, 3,\ldots,n\}$ such that j is a fixed point. Then the intersection of any i of the A_j, for $1 \le i \le n$, has $(n - i)!$ elements, for the $n - i$ not necessarily fixed elements may be permuted arbitrarily. Since there are $\binom{n}{i}$ choices for the A_j that make up the intersection, by the principle of inclusion and exclusion we have

$$|A_1 \cup \cdots \cup A_n| = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} (n-i)!.$$

We conclude that

$$d_n = \sum_{i=0}^{n} (-1)^i \frac{n!}{i!}. \tag{1.18}$$

Inclusion—exclusion principle (probability version). Let $E_1,\ldots, E_n$ be events in a finite sample space. Then

$$\Pr(E_1 \cup \cdots \cup E_n) = \sum_{i=1}^{n} (-1)^{i+1} \sum \Pr(E_{k_1} \cap \cdots \cap E_{k_i}),$$

where the second sum is over all i-tuples $(k_1\ldots,k_i)$ with $1 \le k_1 < \cdots < k_i \le n$.

EXAMPLE 1.24 Stirling numbers of the second kind

Find a formula for the Stirling number of the second kind $\left\{ {n \atop k} \right\}$.

Solution: Using the principle of inclusion and exclusion (see Exercises), we can obtain

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n, \quad 1 \le k \le n.$$

EXAMPLE 1.25 Cards

All 52 playing cards arc dealt randomly to four players, 13 cards per player. What is the probability that at least one person has all cards of the same suit?

Solution: For $1 \le i \le 4$, let E_i be the event that player i has all cards of the same suit. By the principle of inclusion and exclusion, the desired probability, Pr ($\cup$ E_i), is

$$\binom{4}{1}\frac{4}{\binom{52}{13}} - \binom{4}{2}\frac{\binom{4}{2}2!}{\binom{52}{13}\binom{39}{13}} + \binom{4}{3}\frac{\binom{4}{3}3!}{\binom{52}{13}\binom{39}{13}\binom{26}{13}} - \binom{4}{4}\frac{\binom{4}{4}4!}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}$$

$$= 18772910672458601/7450658022984554561005200000$$

$$\doteq 2.5 \times 10^{-11}.$$

Make sure you understand how the four terms in the first line are obtained.

EXAMPLE 1.26 "The problem of derangements"

What is the probability P_n that a random permutation of n elements is a derangement?

Solution: We found in Example 1.23 that

$$d_n = \sum_{i=0}^{n} (-1)^i \frac{n!}{i!}.$$

Therefore

$$P_n = \sum_{i=0}^{n} \frac{(-1)^i}{i!}. \tag{1.19}$$

It may seem strange that a fixed point is less likely to occur when n is 52 than when n is 51 or 53. It is interesting to note that

$$\lim_{n \to \infty} P_n = e^{-1} \doteq 0.37.$$

Students of probability should not be surprised to see the appearance of the number e in a probability calculation.

EXAMPLE 1.27 Average number of fixed points of a permutation

Fine the average number of fixed points of a permutation of n elements.

Solution: We illustrate the result in the case $n = 3$. Below are the permutations of {1, 2, 3} and the number of fixed points of each.

permutation	number of fixed points
(1)(2)(3)	3

(1)(2 3)	1
(2)(1 3)	1
(3)(1 2)	1
(1 2 3)	0
(1 3 2)	0

The permutations an written in cycle form. For instance, the permutation (2)(13) is the one that maps $2 \to 2$, $1 \to 3$, and $3 \to 1$. The total number of fixed points is 6, and the average number is $6/6 = 1$.

Randomly choose a permutation of $\{1, 2, 3, \ldots, n\}$. For $1 \le i \le n$, define $X_i = 1$ if i is fixed and 0 otherwise. Then the number of fixed points is $Y = X_1 + X_2 + \cdots + X_n$. The expected value of each X_i is $(n-1)!/n! = 1/n$. Hence, the expected number of fixed points is

$$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n) = \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = n \cdot \frac{1}{n} = 1.$$

EXAMPLE 1.28 Bell numbers

Find a formula for the Bell number $B(n)$.

Solution: Using the result of Example 1.24, we obtain

$$B(n) = \sum_{k=1}^{n} \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n, \quad n \ge 1.$$

The formula can be simplified considerably:

$$\begin{aligned} B(n) &= \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n \\ &= \sum_{j=0}^{\infty} \frac{j^n}{j!} \sum_{k=j}^{\infty} \frac{(-1)^{k-j}}{(k-j)!} \\ &= \frac{1}{e} \sum_{j=0}^{\infty} \frac{j^n}{j!}. \end{aligned}$$

This formula is interesting from a number-theoretic point of view, as it is not at all clear *a priori* that $(1/e) \sum_{j=0}^{\infty} j^n/j!$ is an integer.

The inclusion–exclusion principle can be generalized to the Bonferroni inequalities of probability theory. We start with the algebraic identity

$$(1+x)^{m-1} = (1+x)^m (1 - x + x^2 - x^3 + \cdots).$$

Equating coefficients of x^k on both sides of this identity yields

$$\binom{m-1}{k} = \sum_{i=0}^{k} \binom{m}{i} (-1)^{k-i}. \tag{1.20}$$

The above identity can also be proved by applying Pascal's identity to the binomial coefficient $\binom{m}{i}$ and collapsing the resulting telescoping sum.

For each $1 \le i \le n$, let

$$N_i = \sum |A_{k_1} \cap \cdots \cap A_{k_i}|, \tag{1.21}$$

where the sum is over all i-tuples $(k_1,\ldots,k_i)$ with $1 \le k_1 < \cdots < k_i \le n$.

Bonferroni inequalities. Let $A_1,\ldots,A_n$ be subsets of a finite set S. If t is an odd number, then

$$|A_1 \cup \cdots \cup A_n| \le \sum_{i=1}^{t} (-1)^{i+1} N_i.$$

If t is even, then the inequality is reversed.

Proof. Let $s \in S$ and assume that s is contained in exactly m of the A_i. If $m = 0$, then the contribution to both sides of the inequality is 0. For $m > 0$, the result follows easily from (1.20).

EXAMPLE 1.29

If k is even, we have

$$d_n \le \sum_{i=0}^{k} (-1)^i \frac{n!}{i!}.$$

If k is odd, then the inequality is reversed.

Here is a neat technique that everyone should learn. Suppose that the sequence

7, 11, 25, 73, 203, 487, 1021, 1925, 3343, 5443, 8417, ...

represents the values of a polynomial $p(n)$, where $n = 0, 1, 2,\ldots$. What is the polynomial?

We take differences of consecutive terms, creating a new sequence:

4, 14, 48, 130, 284, 534, 904, 1418, 2100, 2974, ...

We repeat this process, creating a sequence of sequences:

7,	11,	25,	73,	203,	487,	1021,	1925,	3343,	5443,	8417, ...
4,	14,	48,	130,	284,	534,	904,	1418,	2100,	2974,	...
10,	34,	82,	154,	250,	370,	514,	682,	874,	...	
24,	48,	72,	96,	120,	144,	168,	192,	...		
24,	24,	24,	24,	24,	24,	24,	....			

Having obtained a constant sequence, we stop.

Now, we find the polynomial by multiplying the first column of our difference array by successive binomial coefficients and adding:

$$p(n) = 7\binom{n}{0} + 4\binom{n}{1} + 10\binom{n}{2} + 24\binom{n}{3} + 24\binom{n}{4} = n^4 - 2n^3 + 4n^2 + n + 7.$$

This polynomial gives the original sequence, starting at $p(0)$.

Why does this work? Suppose that the polynomial is

$$p(x) = a_0 + a_1 x + a_2 \frac{x(x-1)}{2!} + \cdots + a_k \frac{x(x-1)\ldots(x-k+1)}{k!},$$

where the a_i are real numbers. A little reflection shows that we can really write an arbitrary polynomial in this way.

Suppose that the values of the polynomial are

$p(0), p(1), p(2), p(3), p(4), p(5), \ldots$.

Letting $n = 0$, we find that $a_0 = p(0)$ (since all the other terms in the polynomial are equal to 0).

The sequence of differences is

$$p(1)-p(0),\ p(2)-p(1),\ p(3)-p(2),\ p(4)-p(3),\ \ldots.$$

Letting $n = 1$, we find that

$$p(1) = a_0 + a_1,$$

and hence

$$a_1 = p(1) - a_0 = p(1) - p(0).$$

The next sequence of differences is

$$p(2)-2p(1)+p(0),\ p(3)-2p(2)+p(1),\ p(4)-2p(3)+p(2),\ p(5)-2p(4)+p(3),\ \ldots.$$

Letting $n = 2$, we obtain

$$p(2) = a_0 + 2a_1 + a_2,$$

and hence

$$a_2 = p(2) - 2p(1) + p(0).$$

This pattern continues, so that the sequence $a_0, a_1, a_2,\ldots$ is the first column of our difference array. In order to establish this, we introduce a little notation. Define

$$\Delta p(n) = p(n+1) - p(n).$$

We call Δ the *difference operator*. We define $\Delta^2 p(n) = \Delta(\Delta p(n))$, $\Delta^3 p(n) = \Delta(\Delta^2 p(n))$, and so on. We have

$$\begin{aligned}
\Delta p(n) &= p(n+1) - p(n)\\
\Delta^2 p(n) &= p(n+2) - 2p(n+1) + p(n)\\
\Delta^3 p(n) &= p(n+3) - 3p(n+2) + 3p(n+1) - p(n)\\
&\vdots\\
\Delta^k p(n) &= \sum_{i=0}^{k} (-1)^{i+k} \binom{k}{i} p(n+i).
\end{aligned}$$

The array of differences looks like

$$\begin{array}{rrrrrrl}
p(0), & p(1), & p(2), & p(3), & p(4), & p(5), & \ldots\\
\Delta p(0), & \Delta p(1), & \Delta p(2), & \Delta p(3), & \Delta p(4), & \Delta p(5), & \ldots\\
\Delta^2 p(0), & \Delta^2 p(1), & \Delta^2 p(2), & \Delta^2 p(3), & \Delta^2 p(4), & \Delta^2 p(5), & \ldots\\
\Delta^3 p(0), & \Delta^3 p(1), & \Delta^3 p(2), & \Delta^3 p(3), & \Delta^3 p(4), & \Delta^3 p(5), & \ldots\\
\Delta^4 p(0), & \Delta^4 p(1), & \Delta^4 p(2), & \Delta^4 p(3), & \Delta^4 p(4), & \Delta^4 p(5), & \ldots\\
\cdots. & & & & & &
\end{array}$$

So our claim is that

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

For a fixed i, the coefficient of $p(i)$ is

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

From the subcommittee identity, we obtain

$$\sum_{k=i}^{n}(-1)^{i+k}\binom{n}{k}\binom{k}{i} = \sum_{k=i}^{n}(-1)^{i+k}\binom{n}{i}\binom{n-i}{k-i}$$

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

$$= (-1)^i\binom{n}{i}\sum_{r=0}^{n-i}(-1)^{i+r}\binom{n-i}{r}$$

$$= \binom{n}{i}\sum_{r=0}^{n-i}(-1)^r\binom{n-i}{r}$$

$$= \begin{cases} 1 & \text{for } n = i \\ 0 & \text{otherwise.} \end{cases}$$

This completes the argument

Theorem. For any n and k, we have

$$\sum_{j=0}^{n}(-1)^{j+k}\binom{n}{j}\binom{j}{k} = \delta(n,k),$$

where $\delta(n, k) = 1$ if $n = k$ and $\delta(n, k) = 0$ if $n \neq k$.

EXERCISES

1.56 (a) Find a formula for the number of surjective (onto) functions, $T(m, n)$, from $\{1, 2, 3, \ldots, m\}$ to $\{1, 2, 3, \ldots, n\}$, where $m \geq n$.

(b) Find a formula for the Stirling number of the second kind $\left\{{n \atop k}\right\}$.

1.57 Euler's ϕ-function is defined as follows:

$$\phi(n) = |\{1 \leq x \leq n : \gcd(x, n) = 1\}|.$$

Find a formula for $\phi(n)$ in terms of the prime factorization of n.

1.58 Prove that

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

if and only if

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

1.59 (*Möbius inversion formula*) (a) Prove that if

$$\sum_{k=1}^{\infty}\alpha(n,k)\beta(k,j) = \sum_{k=1}^{\infty}\beta(n,k)\alpha(k,j) = \delta(n,j),$$

then $f(n) = \sum_{k=1}^{\infty}\alpha(n,k)g(k)$ if and only if $g(n) = \sum_{k=1}^{\infty}\beta(n,k)f(k)$.

(b) Let $\alpha(n, k) = 1$ if $k \mid n$ and 0 otherwise. Determine $\beta(n, k)$.

1.60 What polynomial produces the sequence

1, 4, 13, 34, 73, 136, ...?

1.61 We say that two sets A and B are *linked* if $A \cap B \neq \emptyset$ and neither A nor B is a subset of the other. If S is an n-element set, how many pairs (A, B) of subsets of S exist with A and B linked?

1.7 Fibonacci numbers

Let's discuss one of the most famous sequences of numbers, the Fibonacci sequence. The *Fibonacci sequence* $\{F_0, F_1, F_2, \ldots\}$ is defined recursively by the initial values

(1.22) $F_0 = 0,\ F_1 = 1$

and the recurrence relation

(1.23) $F_n = F_{n-1} + F_{n-2} \quad \text{for } n \geq 2.$

Thus, the Fibonacci numbers are

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144,\ 233,\ 377,\ 610, \ldots.$$

Fibonacci numbers count many things. For example:

- F_{n+1}, is the number of ways that an $n \times 1$ box may be packed with 2×1 and 1×1 boxes.
- F_{n+2} is the number of binary strings of length n that do not contain the substring 00.
- F_{n+2} is the number of subsets of N_n that contain no two consecutive integers.
- F_{n-1} is the number of compositions of n that do not contain 1′s (see Exercise 1.54).

Let's prove the second of these formulas.

Let s_n be the number of binary strings of length n that contain no 00. We will prove that $s_n = F_{n+2}$ for $n \geq 1$. Observe that $s_1 = 2 = F_3$ and $s_2 = 3 = F_4$. We will show that $s_n = s_{n-1} + s_{n-2}$ for $n \geq 3$ (the same recurrence relation satisfied by the Fibonacci numbers). Notice that each binary string of length n that does not contain 00 ends in either 1 or 10. The number of such strings of the first type is s_{n-1} and the number of such strings of the second type is s_{n-2}. Hence $s_n = s_{n-1} + s_{n-2}$ for $n \geq 3$. Now, since $\{s_n\}$ satisfies the same recurrence relation as the Fibonacci numbers, and $s_1 = F_3$ and $s_2 = F_4$, it follows by mathematical induction that $s_n = F_{n+2}$ for all $n \geq 1$.

EXAMPLE 1.30 Cassini's identity

Prove that $F^2_n - F_{n-1} F_{n+1} = (-1)^{n+1}$ for $n \geq 1$.

Solution: We will prove the result by mathematical induction. The identity holds for $n = 1$, since $F^2_1 - F_0F_2 = 1 - 0 = 1 = (-1)^2$. Assume that it holds for n. Then

$$\begin{aligned} F_{n+1}^2 - F_nF_{n+2} &= F_{n+1}^2 - F_n(F_n + F_{n+1}) \\ &= F_{n+1}(F_{n+1} - F_n) - F_n^2 \\ &= F_{n+1}F_{n-1} - F_n^2 \\ &= (-1)^{n+2}. \end{aligned}$$

Hence, the formula holds for $n + 1$ and by induction for all $n \geq 1$.

Here is a delight from Pascal's triangle.

Singmaster's theorem (1975). There are infinitely many numbers that occur at least six times in Pascal's triangle.

Proof. Suppose that we have a solution to

$$r = \binom{n}{m-1} = \binom{n-1}{m}$$

given by

$$m = F_{2k-1}F_{2k}, \quad n = F_{2k}F_{2k+1}, \quad k \geq 2.$$

The number r in such a solution occurs (at least) six times in Pascal's triangle:

$$\binom{r}{1} = \binom{r}{r-1} = \binom{n}{m-1} = \binom{n}{n-m+1} = \binom{n-1}{m} = \binom{n-1}{n-m-1}.$$

The following relations are equivalent:

$$\begin{aligned}
\binom{n}{m-1} &= \binom{n-1}{m} \\
\frac{n!}{(m-1)!(n-m+1)!} &= \frac{(n-1)!}{m!(n-m-1)!} \\
nm &= (n-m+1)(n-m) \\
F_{2k-1}F_{2k}F_{2k}F_{2k+1} &= (F_{2k}F_{2k+1} - F_{2k-1}F_{2k} + 1)(F_{2k}F_{2k+1} - F_{2k-1}F_{2k}) \\
&= [F_{2k}(F_{2k+1} - F_{2k-1}) + 1][F_{2k}(F_{2k+1} - F_{2k-1})] \\
&= (F_{2k}^2 + 1)F_{2k}^2 \\
F_{2k-1}F_{2k+1} &= F_{2k}^2 + 1.
\end{aligned}$$

The last relation is true by Cassini's identity.

The smallest such number given by our proof (when $k = 2$) is 3003.

EXERCISES

1.62 Prove the identity

$$F_1 + \cdots + F_n = F_{n+2} - 1, \quad n \geq 1.$$

1.63 Prove the identity

$$F_1^2 + \cdots + F_n^2 = F_nF_{n+1}, \quad n \geq 1.$$

1.64 Prove the identity

$$F_{m+n} = F_mF_{n+1} + F_{m-1}F_n, \quad m \geq 1, n \geq 0.$$

1.65 Where do you find Fibonacci numbers in Pascal's triangle? What identity proves this?

1.66 Find positive integers n and k, with $k < n$, for which

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n}{k+2}.$$

1.67 Prove that

$$\sum_{n=1}^{\infty} \tan^{-1} \frac{1}{F_{2n+1}} = \frac{\pi}{4}.$$

1.68 Find the least number greater than 1 that occurs six times in Pascal's triangle.

1.8 Linear recurrence relations

A sequence $\{a_n\}$ satisfies a *linear homogeneous recurrence relation with constant coefficients* (of order k) if

$$a_n = \sum_{i=1}^{k} c_i a_{n-i} \tag{1.24}$$

for constants $c_1,\ldots,c_k$ and all $n \geq k$.

The Fibonacci sequence $\{F_n\}$ satisfies a linear homogeneous recurrence relation with constant coefficients (of order 2).

How fast do the Fibonacci numbers grow? One might guess that they grow exponentially and they do. In order to find the exact rate of growth, we first find an explicit formula for F_n.

We will show how to guess and construct a solution. Assume that x^n, where $n \geq 0$, is the general term of a sequence that satisfies the Fibonacci recurrence relation (but not necessarily the same initial conditions). Then

$$x^n = x^{n-1} + x^{n-2}.$$

Assuming that $x \neq 0$, we divide through by x^n and obtain the equation

$$x^2 - x - 1 = 0. \tag{1.25}$$

This polynomial $x^2 - x - 1$ is called the *characteristic polynomial* of the sequence. We use the quadratic formula to find the two roots of the characteristic polynomial:

$$\phi = \frac{1+\sqrt{5}}{2}, \quad \hat{\phi} = \frac{1-\sqrt{5}}{2}. \tag{1.26}$$

We call ϕ the "golden ratio." Note that $\phi \doteq 1.6$ and $\hat{\phi} \doteq -0.6$.

So we know that ϕ^n and $\hat{\phi}^n$ both satisfy the Fibonacci recurrence relation. Any linear combination $A\phi^n + B\hat{\phi}^n$, with $A, B \in \mathbf{R}$, also satisfies the recurrence relation. For

$$\begin{aligned}(A\phi^{n-1} + B\hat{\phi}^{n-1}) + (A\phi^{n-2} + B\hat{\phi}^{n-2}) &= A(\phi^{n-1} + \phi^{n-2}) + B(\hat{\phi}^{n-1} + \hat{\phi}^{n-2}) \\ &= A\phi^n + B\hat{\phi}^n.\end{aligned}$$

We use the initial conditions to solve for the coefficients A and B. Recalling that $F_0 = 1$ and $F_1 = 1$, we obtain two linear equations to solve simultaneously:

$$1 = A\phi^0 + B\hat{\phi}^0 = A + B$$

$$1 = A\phi^1 + B\hat{\phi}^1 = A\left(\frac{1+\sqrt{5}}{2}\right) + B\left(\frac{1-\sqrt{5}}{2}\right).$$

We find that

$$A = \frac{1}{\sqrt{5}} \text{ and } B = -\frac{1}{\sqrt{5}}.$$

Thus, a formula for the Fibonacci numbers is

$$F_n = \frac{\phi^n - \hat{\phi}^n}{\sqrt{5}}, \quad n \geq 0. \tag{1.27}$$

The above function satisfies the recurrence relation and initial conditions of the Fibonacci sequence, and hence is a formula for the Fibonacci sequence (since the sequence is well-defined). But the derivation of the formula was based on the assumption that some basic solutions to the recurrence relation were exponential. How did we know this in advance and would it be true for other linear recurrence relations? A more direct way to solve these problems is via generating functions, which we address in Chapter 2.

Now, how do we evaluate the growth rate of F_n? We say that a positive-valued function $f(n)$ is *asymptotic* to another such function $g(n)$, and we write $f(n) \sim g(n)$, if $\lim_{n\to\infty} f(n)/g(n) = 1$. Since $\hat{\phi}^n \to 0$ as $n \to \infty$, we conclude that

$$F_n \sim \frac{\phi^n}{\sqrt{5}}. \tag{1.28}$$

EXAMPLE 1.31

Find an explicit formula for the sequence $\{a_n\}$ defined by the recurrence formula

$$a_0 = 1,\ a_1 = 1,\ a_n = 6a_{n-1} - 9a_{n-2},\ n \geq 2.$$

Solution: The characteristic polynomial of the sequence is

$$x^2 - 6x + 9 = (x-3)^2,$$

which has 3 as a double root. Hence, 3^n is a solution to the recurrence relation. However, we need a second solution in order to make the formula satisfy the initial conditions. A guess for a second solution is $n3^n$. Let's verify that this solution satisfies the recurrence relation:

$$\begin{aligned} 6(n-1)3^{n-1} - 9(n-2)3^{n-2} &= 3^{n-2}(18n - 18 - 9n + 18) \\ &= n3^n. \end{aligned}$$

Any linear combination of our two solutions also satisfies the recurrence relation:

$$A3^n + Bn3^n.$$

In order to satisfy the initial values, $a_0 = 1$ and $a_1 = 1$, we require that

$$1 = A$$
$$1 = 3A + 3B,$$

and hence $A = 1$ and $B = -2/3$. Therefore, an explicit formula for the sequence is

$$a_n = 3^n - 2n3^{n-1}, \quad n \geq 0.$$

The next example illustrates the technique of adding a particular solution and a homogeneous solution.

EXAMPLE 1.32

Find an explicit formula for the sequence $\{a_n\}$ defined by the recurrence formula

$$a_0 = 1,\ a_1 = 1,\ a_n = 6a_{n-1} - 9a_{n-2} + n,\ n \geq 2.$$

Solution: We find a particular solution to the recurrence relation. Assume the existence of a solution of the form $a_n = \alpha n + \beta$, where α and β are constants. Thus

$$\alpha n + \beta = 6(\alpha(n-1) + \beta) - 9(\alpha(n-2) + \beta) + n$$
$$(4\alpha - 1)n = 12\alpha - 4\beta.$$

In order for this identity to hold for all n, we must have $\alpha = 1/4$ and hence $\beta = 3/4$.

Therefore

$$\frac{1}{4}n + \frac{3}{4}$$

satisfies the recurrence relation.

We solved the homogeneous version of this recurrence relation in the previous example. Thus, the general solution to the recurrence relation is of the form

$$A3^n + Bn3^n + \frac{1}{4}n + \frac{3}{4}.$$

The initial values, $a_0 = 1$ and $a_1 = 1$, determine the values $A = 1/4$ and $B = -1/4$.

Therefore, an explicit formula is

$$a_n = \frac{1}{4}3^n - \frac{1}{4}n3^n + \frac{1}{4}n + \frac{3}{4}, \quad n \geq 0.$$

The *Lucas numbers* are defined as

(1.29) $$L_0 = 2,\ L_1 = 1,\ L_n = L_{n-1} + L_{n-2}, \quad n \geq 2.$$

Thus, the Lucas numbers are

$$2,\ 1,\ 3,\ 4,\ 7,\ 11,\ 18,\ 29,\ 47,\ 76,\ 123,\ 199,\ 322,\ 521,\ 843,\ 1364,\ \ldots.$$

- L_n is the number of ways that an $n \times 1$ box may be packed with 2×1 and 1×1 boxes, allowing "wrap-around."
- L_n is the number of subsets of $\{1,\ldots,n\}$ which do not contain two consecutive numbers, where 1 and n are considered consecutive.

Since the Lucas numbers satisfy the same recurrence relation as the Fibonacci numbers, they have the same characteristic polynomial, $x^2 - x - 1$. Taking into account the initial values $L_0 = 2$ and $L_1 = 1$, we obtain a formula for the Lucas numbers:

(1.30) $$L_n = \phi^n + \hat{\phi}^n, \quad n \geq 0.$$

The simplicity of this formula is one of the nice properties of the Lucas sequence. A consequence is that the Lucas numbers are given by the elegant formula $L_n = \{\phi^n\}$ for $n \geq 2$, where $\{x\}$ is the nearest integer to x.

EXAMPLE 1.33 Squares of Fibonacci numbers

Let $\{F^2_n\}$ be the sequence of squares of the Fibonacci numbers. Find a linear recurrence relation with constant coefficients for this sequence.

Solution: Start with the relations

$$F_n = F_{n-1} + F_{n-2}$$
$$F_{n-3} = F_{n-1} - F_{n-2}.$$

Square both relations and add:

$$\begin{aligned} F_n^2 + F_{n-3}^2 &= (F_{n-1} + F_{n-2})^2 + (F_{n-1} - F_{n-2})^2 \\ &= 2F_{n-1}^2 + 2F_{n-2}^2. \end{aligned}$$

We obtain the recurrence relation

$$F_n^2 = 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2, \quad n \geq 3.$$

EXAMPLE 1.34 Powers of Fibonacci numbers

Find a linear recurrence relation with constant coefficients for the sequence $\{F^k{}_n\}$ of kth powers of the Fibonacci numbers, where k is a positive integer.

Solution: The method is to use characteristic polynomials. Let's work out the $k = 2$ case first (this will reproduce the recurrence relation found in the previous example). We know a direct formula for the Fibonacci numbers:

$$F_n = A\phi^n + B\hat{\phi}^n, \quad n \geq 0,$$

where A and B are constants (we know the constants but don't need them). It follows that

$$F_n^2 = A^2(\phi^2)^n + 2AB(\phi\hat{\phi})^n + B^2(\hat{\phi}^2)^n,$$

and since $\phi\hat{\phi} = -1$, the roots of the characteristic polynomial for the sequence $\{F_n^2\}$ are ϕ^2, $\hat{\phi}^2$, and -1. Hence, the characteristic polynomial for this sequence is

$$(x - \phi^2)(x - \hat{\phi}^2)(x + 1) = [x^2 - (\phi^2 + \hat{\phi}^2)x + 1](x + 1).$$

To simplify further, recall that the Lucas numbers L_n are given by the formula

$$L_n = \phi^n + \hat{\phi}^n, \quad n \geq 0.$$

Using this formula, the characteristic polynomial in the case $k = 2$ simplifies to

$$(x^2 - L_2x + 1)(x + 1) = (x^2 - 3x + 1)(x + 1) = x^3 - 2x^2 - 2x + 1.$$

This confirms the recurrence relation found in the previous example:

$$F_n^2 = 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2, \quad n \geq 3.$$

The case $k = 3$ is similar. By the binomial theorem, the formula for $F^3{}_n$ contains powers of ϕ^3, $\phi^2\hat{\phi} = -\phi$, $\phi\hat{\phi}^2 = -\hat{\phi}$, and $\hat{\phi}^3$. Therefore, the characteristic polynomial of the sequence $\{F_n^3\}$ is

$$\begin{aligned}(x - \phi^3)(x - \hat{\phi}^3)(x + \phi)(x + \hat{\phi}) &= [x^2 - (\phi^3 + \hat{\phi}^3)x - 1][x^2 + (\phi + \hat{\phi})x - 1] \\ &= (x^2 - L_3x - 1)(x^2 + L_1x - 1) \\ &= (x^2 - 4x - 1)(x^2 + x - 1) \\ &= x^4 - 3x^3 - 6x^2 + 3x + 1.\end{aligned}$$

This gives us a recurrence relation for the cubes of the Fibonacci numbers:

$$F_n^3 = 3F_{n-1}^3 + 6F_{n-2}^3 - 3F_{n-3}^3 - F_{n-4}^3, \quad n \geq 4.$$

For $k \geq 1$, the characteristic polynomial for the sequence of kth powers of the Fibonacci numbers is

$$\prod_{i=0}^{\lfloor \frac{k-1}{2} \rfloor} [x^2 + (-1)^{i+1}L_{k-2i}x + (-1)^k] \cdot \begin{cases} 1 & \text{if } k \bmod 4 = 1, 3 \\ (x - 1) & \text{if } k \bmod 4 = 0 \\ (x + 1) & \text{if } k \bmod 4 = 2 \end{cases}.$$

This formula, found by John Riordan, means that the sequence $\{F^k{}_n\}$ of kth powers of the Fibonacci numbers satisfies a linear recurrence relation of order $k + 1$ with integer coefficients.

EXERCISES

1.69 Let $\{a_n\}$ be defined by the recurrence

$$a_0 = 0,\ a_1 = 1,\ a_n = 5a_{n-1} - 6a_{n-2},\ n \geq 2.$$

Find an explicit formula for a_n.

1.70 Suppose that the sequence $\{a_n\}$ satisfies the recurrence relation

$$a_n = 3a_{n-1} + 4a_{n-2} - 12a_{n-3}, \quad n \geq 3,$$

where $a_0 = 0$, $a_1 = 1$, and $a_2 = 2$. Find an explicit formula for a_n.

1.71 Let $\{b_n\}$ be defined by the recurrence

$$b_0 = 0,\ b_1 = 0,\ b_2 = 1,\ b_n = 4b_{n-1} - b_{n-2} - 6b_{n-3},\ n \geq 3.$$

Find an explicit formula for b_n.

Do the same where the initial values are $b_0 = 0$, $b_1 = 1$, $b_2 = 2$.

1.72 Define $\{a_n\}$ by the recurrence

$$a_0 = 0,\ a_1 = 1,\ a_n = 5a_{n-1} - 6a_{n-2},\ n \geq 2$$

and $\{b_n\}$ by the recurrence

$$b_0 = 0,\ b_1 = 1,\ b_n = 9b_{n-1} - 20b_{n-2},\ n \geq 2.$$

Find a linear recurrence for the sequence $\{c_n\}$ defined by

$$c_n = a_n + b_n, \quad n \geq 0.$$

Find a linear recurrence for the sequence $\{d_n\}$ defined by

$$d_n = a_n b_n, \quad n \geq 0.$$

1.73 (a) Find a recurrence formula for the sequence $\{a_n\}$ defined by $a_n = 3^n + n^2$, where $n \geq 0$.

(b) Find a recurrence formula for the sequence $\{a_n\}$ defined by $a_n = 3^n + n^2 + 6n + 7$, where $n \geq 0$.

1.74 Find an explicit formula for the sequence $\{a_n\}$ defined by the recurrence relation

$$a_0 = 0,\ a_1 = 1,\ a_n = a_{n-1} + a_{n-2} + n,\ n \geq 2.$$

1.75 Find an explicit formula for the sequence $\{a_n\}$ defined by the recurrence relation

$$a_0 = 1,\ a_1 = 3,\ a_n = a_{n-1} + a_{n-2} + 2^{n-2},\ n \geq 2.$$

How fast does a_n grow?

1.76 Prove the identity $L_n = F_{n-1} + F_{n+1}$ for $n \geq 1$.

1.77 Prove the identity $F_n = (L_{n-1} + L_{n+1})/5$ for $n \geq 1$.

1.78 Prove the identity $F_{2n} = F_n L_n$ for $n \geq 0$.

1.79 Prove the identity $L_{3n} = L^3_n - 3(-1)^n L_n$ for $n \geq 0$.

1.80 Find a linear recurrence relation satisfied by all cubic polynomials.

1.81 Find a linear homogeneous recurrence relation (not with constant coefficients) for the sequence $\{a_n\}$, where $a_n = 2^n + n!$.

1.82 A *square number* is a number of the form n^2, where n is a nonnegative integer. A *triangular number* is a number of the form $n(n + 1)/2$, where n is a nonnegative integer. Let a_n be the nth number that is both square and triangular. For example, $a_0 = 0$, $a_1 = 1$, and $a_2 = 36$. Find a linear homogeneous recurrence relation with constant coefficients for $\{a_n\}$.

1.83 Suppose that $a_n = c_0 + \Sigma^k_{i=1} c_i a_{n-i}$, for $n \geq k$. Prove that $\{a_n\}$ satisfies a linear homogeneous recurrence relation (with constant coefficients) of order $k + 1$.

1.9 Special recurrence relations

Recall that a derangement is a permutation with no fixed points. Let d_n denote the number of derangements of the set $\{1, 2, 3,\ldots,n\}$. We know that $d_1 = 0$ and $d_2 = 1$. We claim that $\{d_n\}$ satisfies the linear recurrence relation

$$d_n = (n-1)(d_{n-1} + d_{n-2}), \quad n \geq 3. \tag{1.31}$$

In a derangement of $\{1, 2, 3,\ldots,n\}$, the element n must occur in a cycle of length 2 or a cycle of greater length. There are $n-1$ choices for the other element in a cycle of length 2, and the remaining elements constitute a derangement of $n-2$ elements. In a cycle of length greater than 2, there are $n-1$ choices for the element which maps to n, and the elements other than n constitute a derangement of $n-1$ elements.

We defined the Stirling number of the second kind $\left\{ {n \atop k} \right\}$, for $1 \leq k \leq n$, to be the number of partitions of the set $\{1, 2, 3,\ldots,n\}$ into k parts. We will now find a recurrence formula for these numbers. Note that $\left\{ {n \atop 1} \right\} = 1$ for all n, as there is only one way to partition $\{1, 2, 3,\ldots,n\}$ into one subset. Also, $\{{}^{n}_{n}\} = 1$ for all n, as $\{1, 2, 3,\ldots,n\}$ may be partitioned into n subsets in only one way. Now let us find a way to compute $\{{}^{n}_{k}\}$ from previous values. In a partition of $\{1, 2, 3,\ldots,n\}$ into k parts, the element n can be alone in a part of the partition or it can be in a part with other elements. If it is alone, then there are $\{{}^{n-1}_{k-1}\}$ ways to partition $\{1, 2,\ldots,n-1\}$ into the other $k-1$ parts. However, if n is in a part with other elements, then there are k choices for which part contains n and $\{{}^{n-1}_{k}\}$ ways to partition $\{1, 2, \ldots,n-1\}$ into k parts. Therefore

$$\left\{ {n \atop 1} \right\} = 1 \quad \text{for } n \geq 1$$

$$\left\{ {n \atop n} \right\} = 1 \quad \text{for } n \geq 1$$

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\} \quad \text{for } 2 \leq k < n. \tag{1.32}$$

From the recurrence formula, we obtain a table (Table 1.1) of values of $\{{}^{n}_{k}\}$ for small n and k. The row sums of this table are the Bell numbers.

Table 1.1 Stirling numbers of the second kind $\{{}^{n}_{k}\}$ and Bell numbers $B(n)$.

n \ k	1	2	3	4	5	6	7	$B(n)$
1	1							1
2	1	1						2
3	1	3	1					5
4	1	7	6	1				15
5	1	15	25	10	1			52
6	1	31	90	65	15	1		203
7	1	63	301	350	140	21	1	877

Let us verify an entry of the table, say $\{^4_3\} = 6$. There are six ways to partition the set $\{1, 2, 3, 4\}$ into three subsets: $\{12, 3, 4\}$, $\{1, 3, 24\}$, $\{1, 2, 34\}$, $\{13, 2, 4\}$, $\{1, 4, 23\}$, $\{14, 2, 3\}$ (suppressing commas and one level of set notation).

The *Stirling number of the first kind* $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$, for $1 \le k \le n$, is defined to be the number of permutations of $\{1, 2, 3, \ldots, n\}$ that have k cycles. For example, $\left[\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right] = 3$, as there are three permutations of $\{1, 2, 3\}$ with two cycles: (1 2)(3), (1 3)(2), and (2 3)(1).

Observe that $\left[\begin{smallmatrix} n \\ n \end{smallmatrix}\right] = 1$ (there is only one identity permutation) and $\left[\begin{smallmatrix} n \\ 1 \end{smallmatrix}\right] = (n-1)!$ (there are $(n-1)!$ ways to seat n guests at a circular table). In a permutation of $\{1, \ldots, n\}$, the element n can constitute a cycle by itself or it can follow one of the other $n-1$ elements in one of k cycles. In the first case, there are $\left[\begin{smallmatrix} n-1 \\ k-1 \end{smallmatrix}\right]$ choices for dividing the other $n-1$ elements into $k-1$ cycles. In the second case, there are $n-1$ choices for which element n follows and $\left[\begin{smallmatrix} n-1 \\ k \end{smallmatrix}\right]$ ways to divide $n-1$ elements into k cycles. Therefore

$$\begin{bmatrix} n \\ 1 \end{bmatrix} = (n-1)! \quad \text{for } n \ge 1$$

$$\begin{bmatrix} n \\ n \end{bmatrix} = 1 \quad \text{for } n \ge 2$$

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} + (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix} \quad \text{for } 2 \le k < n. \tag{1.33}$$

From this recurrence formula, we obtain a table (Table 1.2) of the values of $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ for small n and k. Note that the sum of the entries of the nth row of the table is $n!$, which is correct because each permutation of $\{1, 2, 3, \ldots, n\}$ is counted.

Table 1.2 Stirling numbers of the first kind $[^n_k]$.

n \ k	1	2	3	4	5	6	7
1	1						
2	1	1					
3	2	3	1				
4	6	11	6	1			
5	24	50	35	10	1		
6	120	274	225	85	15	1	
7	720	1764	1624	735	175	21	1

We set $\left\{\begin{smallmatrix} n \\ k \end{smallmatrix}\right\} = \left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right] = 0$ for $k > n$ or $k = 0$, and $\left\{\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right\} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] = 1$. Stirling numbers of the first and second kinds are linked by a simple identity:

$$\begin{Bmatrix} n \\ k \end{Bmatrix} = \begin{bmatrix} -k \\ -n \end{bmatrix} \quad \text{for all } k, n. \tag{1.34}$$

This means that the Stirling numbers are represented in dovetailing arrays (Table 1.3).

Table 1.3 Stirling numbers of the first and second kinds.

n \ k	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5	1										
−4	10	1									
−3	35	6	1								
−2	50	11	3	1							
−1	24	6	2	1	1						
0	0	0	0	0	0	1					
1	0	0	0	0	0	0	1				
2	0	0	0	0	0	0	1	1			
3	0	0	0	0	0	0	1	3	1		
4	0	0	0	0	0	0	1	7	6	1	
5	0	0	0	0	0	0	1	15	25	10	1

Recall that $p(n)$ is the number of partitions of n units into an arbitrary number of parts, while $p(n, k)$ is the number of partitions of n units into k parts. Clearly, $p(n) = \sum_{k=1}^{n} p(n, k)$.

A recurrence relation formula calculates $p(n, k)$:

$$p(1,1) = 1$$
$$p(n,k) = 0 \quad k > n \text{ or } k = 0,$$
$$p(n,k) = p(n-1,k-1) + p(n-k,k) \quad n \geq 2 \text{ and } 1 \leq k \leq n.$$

(1.35)

The value $p(1, 1) = 1$ is obvious. Since there are no partitions of n into more than n parts or into 0 parts, we have $p(n, k) = 0$ for $k > n$ or $k = 0$. In a partition of n into k parts, the smallest part is either 1 or greater than 1. In the former case, there are $p(n - 1, k - 1)$ partitions of the remaining number $n - 1$ into $k - 1$ parts. In the latter case, the partitions of n into k parts are equinumerous with the partitions of $n - k$ into k parts (just subtract 1 from each part in the partition of n). This proves the formula $p(n, k) = p(n - 1, k - 1) + p(n - k, k)$ for $n \geq 2$ and $1 \leq k \leq n$.

Tables 1.4 and 1.5 show values of $p(n, k)$ and $p(n)$ for small n and k.

Table 1.4 Partition numbers $p(n, k)$.

n \ k	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	1	1							
4	1	2	1	1						
5	1	2	2	1	1					
6	1	3	3	2	1	1				
7	1	3	4	3	2	1	1			
8	1	4	5	5	3	2	1	1		
9	1	4	7	6	5	3	2	1	1	
10	1	5	8	9	7	5	3	2	1	1

Table 1.5 Partition numbers $p(n)$.

n	$p(n)$	n	$p(n)$	n	$p(n)$	n	$p(n)$
1	1	26	2436	51	239943	76	9289091
2	2	27	3010	52	281589	77	10619863
3	3	28	3718	53	329931	78	12132164
4	5	29	4565	54	386155	79	13848650
5	7	30	5604	55	451276	80	15796476
6	11	31	6842	56	526823	81	18004327
7	15	32	8349	57	614154	82	20506255
8	22	33	10143	58	715220	83	23338469
9	30	34	12310	59	831820	84	26543660
10	42	35	14883	60	966467	85	30167357
11	56	36	17977	61	1121505	86	34262962
12	77	37	21637	62	1300156	87	38887673
13	101	38	26015	63	1505499	88	44108109
14	135	39	31185	64	1741630	89	49995925
15	176	40	37338	65	2012558	90	56634173
16	231	41	44583	66	2323520	91	64112359
17	297	42	53174	67	2679689	92	72533807
18	385	43	63261	68	3087735	93	82010177
19	490	44	75175	69	3554345	94	92669720
20	627	45	89134	70	4087968	95	104651419
21	792	46	105558	71	4697205	96	118114304
22	1002	47	124754	72	5392783	97	133230930
23	1255	48	147273	73	6185689	98	150198136
24	1575	49	173525	74	7089500	99	169229875
25	1958	50	204226	75	8118264	100	190569292

EXERCISES

1.84 Use a computer to calculate d_{100}.

1.85 Prove the formula $d_n = nd_{n-1} + (-1)^n$ for $n \geq 2$.

1.86 Find (with proof) a formula for $[_{n-1}{}^n]$ for $n \geq 2$.

1.87 For $n \geq 2$, prove that among the permutations of an n-element set, there are as many with an even number of disjoint cycles as with an odd number of disjoint cycles. This explains why the alternate addition and subtraction of the entries of any row n, with $n \geq 2$, of the table of Stirling numbers of the first kind is equal to 0.

1.88 Show that the Bell numbers $B(n)$ satisfy the recurrence formula

$$B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(k), \quad n \geq 0,$$

where $B(0) = 1$.

1.89 Prove that the expected number of parts in a random partition of $\{1,2,3,\ldots,n\}$ is $(B(n+1) - B(n))/B(n)$, where $B(n)$ is the nth Bell number.

1.90 Show that the recurrence relations for the Stirling numbers of the first and second kinds (allowing for negative values of the arguments) are equivalent.

1.91 Show that $p(n, k) = \Sigma^k_{j=1} p(n-k, j)$.

1.92 Let b_n be the number of order-preserving labelings of the complete binary tree with $2^n - 1$ nodes using the integers $\{1, 2,\ldots,2^n - 1\}$. Show that $b_1 = 1$ and $b_n = b_{n-1}^2 \binom{2^n-2}{2^{n-1}-1}$ for $n \geq 2$.

1.10 Counting and number theory

In this section, we investigate divisibility properties of factorials, binomial coefficients, and Fibonacci numbers.

EXAMPLE 1.35

How many 0′s occur at the right of 40!?

Solution: The 0′s at the right of 40! occur because of factors of 2 and 5 among the numbers 1. 2,…,40. Since there are more 2′s than 5′s, the number of 0′s is determined by the exponent of 5 that divides 40!. This number is

$$\sum_{k=1}^{\infty} \left\lfloor \frac{40}{5^k} \right\rfloor = \left\lfloor \frac{40}{5} \right\rfloor + \left\lfloor \frac{40}{25} \right\rfloor = 8 + 1 = 9.$$

In the following discussion, let p be a prime.

De Polignac's formula. The exponent to which p divides $n!$ is given by

$$\sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor.$$

Let $d_b(n)$ be the sum of the "digits" in the base b representation of n. For instance, if the base 3 representation of n is 1020120, then $d_3(n) = 6$.

Theorem. The exponent of 2 that divides $n!$ is $n - d_2(n)$.

Proof. Let the base 2 representation of n be

$$n = b_k b_{k-1} \cdots b_1 b_0.$$

Then $n = \Sigma^k_{i=0} b_i 2^i$ and the exponent of 2 that divides $n!$ is

$$\begin{aligned}\sum_{i=1}^{\infty} \left\lfloor \frac{n}{2^i} \right\rfloor =& (b_1 + 2b_2 + 2^2 b_3 + \cdots + 2^{k-1} b_k) \\ &+ (b_2 + 2b_3 + \cdots + 2^{k-2} b_k) \\ &+ \cdots \\ &+ b_k \\ =& (2^0 - 1)b_0 + (2^1 - 1)b_1 + (2^2 - 1)b_2 + \cdots + (2^k - 1)b_k \\ =& n - d_2(n).\end{aligned}$$

Here is the general version of the theorem.

Legendre's formula (1808). The exponent of p that divides $n!$ is

$$\frac{n - d_p(n)}{p - 1}.$$

Next we will look at divisibility of binomial coefficients.

Theorem. If $1 \le k \le p - 1$, then $\binom{p}{k}$ is divisible by p.

Proof. The numerator of $p!/(k!(p-k)!)$ is a multiple of p and p does not divide the denominator.

Kummer's theorem (1852). The exponent to which p divides the binomial coefficient $\binom{n}{k}$ is equal to the number of carries when k and $n - k$ are added in base p.

Proof. We will show the proof in the base 2 case. Let $j = n - k$. The exponent to which 2 divides $\binom{n}{k}$ is

$$n - d_2(n) - (j - d_2(j) + k - d_2(k)) = d_2(j) + d_2(k) - d_2(n).$$

Assume that the binary representation of n requires l binary digits. For $1 \le i \le l$, let n_i, j_i and k_i be the ith binary digit of the expansion of n, j, and k, respectively; let $c_i = 1$ if there is a carry in the ith place when j and k are added (in binary) and $c_i = 0$ if there is no carry. Also, define $c_{-1} = 0$. We see that n_i

$= j_i + k_i + c_{i-1} - 2c_i$ for $1 \le i \le l$. Hence, the exponent to which 2 divides $\binom{n}{k}$ is

$$\sum_{i=0}^{l}(j_i + k_i - n_i) = \sum_{i=0}^{l}(2c_i - c_{i-1}) = \sum_{i=0}^{l} c_i.$$

Corollary. For $e \ge 1$ and $1 \le x < p^e$, we have

$$\binom{p^e}{x} \equiv 0 \pmod{p}.$$

The next theorem gives a practical method for calculating binomial coefficients modulo p.

Lucas' theorem (1878). Suppose that $0 \le a_i, b_i < p$ for $1 \le i \le k$. Then

$$\binom{a_0 + a_1 p + a_2 p^2 + \cdots + a_k p^k}{b_0 + b_1 p + b_2 p^2 + \cdots + b_k p^k} \equiv \binom{a_0}{b_0}\binom{a_1}{b_1}\binom{a_2}{b_2}\cdots\binom{a_k}{b_k} \pmod{p}.$$

Proof. The left side counts the ways of choosing $b_0 + b_{1p} + b_2 p^2 + \cdots + b_k p^k$ balls from a set of $a_0 + a_1 p + a_2 p^2 + \cdots + a_k p^k$ balls. Suppose that the balls to be selected are in boxes, with b_0 boxes containing a single ball each, b_1 boxes containing p balls each, b_2 boxes containing p^2 balls each, …,and b_k boxes containing p^k balls each. In selecting the balls from the boxes, any choice of only some balls from a box leads to a contribution of 0 (mod p), since $\binom{p^e}{x} \equiv 0 \pmod{p}$ for $1 \le x < p^e$. Hence, the only selections that matter (modulo p) are those that take none or all the balls from a particular box. This means that we need to select b_i boxes from a set of a_i boxes from which to take all the balls, for $0 \le i \le k$. The number of ways to do this is

$$\binom{a_0}{b_0}\binom{a_1}{b_1}\binom{a_2}{b_2}\cdots\binom{a_k}{b_k}.$$

Say that the base b representation of m *dominates* the base b representation of n if the former is greater than the latter in each place.

Corollary. The binomial coefficient $\binom{n}{k}$ is divisible by p if and only if the base p representation of n does not dominate the base p representation of k.

EXAMPLE 1.36

Is $\binom{59}{12}$ divisible by 7?

Solution: We have $59 = 1 \cdot 7^2 + 1 > + 3$ and $12 = 1 \cdot 7 + 5$. Since the base 7 representation of 59 does not dominate the base 7 representation of 12, we conclude that $\binom{59}{12}$ is divisible by 7.

Here is a charming result about numbers in a row of Pascal's triangle.

Erdős and Szekeres theorem (1978). In a row of Pascal's triangle, any two numbers other than the l's have a common factor.

Proof. Suppose that the numbers are $\binom{n}{j}$ and $\binom{n}{k}$, with $0 < j < k < n$. Then, by the subcommittee identity.

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

Obviously, $\binom{n}{j}$ divides the right side of this relation, so it also divides the left side.

However, if $\binom{n}{j}$ and $\binom{n}{k}$ were coprime, then $\binom{n}{j}$ would divide $\binom{k}{j}$, but this is impossible since $\binom{n}{j} > \binom{k}{j}$.

It is not known whether there are infinitely many Fibonacci numbers that are primes. There are infinitely many composite Fibonacci numbers, since every third Fibonacci number is even. We will show that there exist relatively prime positive integers a and b such that the Fibonacci-like sequence defined by the Fibonacci recurrence relation and the initial values a and b contains *no* prime numbers. Our sequence $\{a_n\}$ is defined by

$$a_0 = a,\ a_1 = b,\ a_n = a_{n-1} + a_{n-2},\ n \geq 2.$$

It follows by mathematical induction that

$$a_n = aF_{n-1} + bF_n, \quad n \geq 1.$$

We define 17 quadruples of integers (p_i, m_i, r_i, c_i), where $1 \leq i \leq 17$. These quadruples satisfy the following properties:

(1) each p_i is prime;

(2)$p_i | F_{m_i}$;

(3) the congruences $x \equiv r_i \pmod{m_i}$ cover all the integers, that is, given any integer n, one of the congruences is satisfied by n.

The purpose of the c_i is to control the size of a and b.

We define

$$a \equiv c_i F_{m_i - r_i} \pmod{p_i}, \quad b \equiv c_i F_{m_i - r_i + 1} \pmod{p_i}, \quad \text{for all } i.$$

Such a and b exist by the Chinese remainder theorem.

Chinese Remainder Theorem. If $n_1, n_2, \ldots, n_k$ are pairwise relatively prime numbers, and $r_1, r_2, \ldots, r_k$ are any integers, then there exists an integer x satisfying the simultaneous congruences

$$\begin{aligned} x &\equiv r_1 \pmod{n_1} \\ x &\equiv r_2 \pmod{n_2} \\ &\vdots \\ x &\equiv r_k \pmod{n_k}. \end{aligned}$$

Furthermore, x is unique modulo $n_1 n_2 \ldots n_k$.

It follows that

$$\begin{aligned} a_n &\equiv c_i F_{m_i - r_i} F_{n-1} + c_i F_{m_i - r_i + 1} F_n \pmod{p_i} \\ &\equiv c_i (F_{m_i - r_i} F_{n-1} + F_{m_i - r_i + 1} F_n) \pmod{p_i} \\ &\equiv c_i F_{m_i - r_i + n} \pmod{p_i}. \end{aligned}$$

Since $F_m \mid F_{mn}$, we have $p_i \mid a_n$.

The following collection of 17 quadruples was found by Maxim Vsemirnov:

$(3, 4, 3, 2)$	$(2, 3, 1, 1)$	$(5, 5, 4, 2)$
$(7, 8, 5, 3)$	$(17, 9, 2, 5)$	$(11, 10, 6, 6)$
$(47, 16, 9, 34)$	$(19, 18, 14, 14)$	$(61, 15, 12, 29)$
$(23, 24, 17, 6)$	$(107, 36, 8, 19)$	$(31, 30, 0, 21)$
$(1103, 48, 33, 9)$	$(181, 90, 80, 58)$	$(41, 20, 18, 11)$
	$(541, 90, 62, 85)$	$(2521, 60, 48, 306)$

Using the Chinese remainder theorem, we find

$$a = 106276436867,\ b = 35256392432.$$

These are composite numbers with no common factor.

EXERCISES

1.93 How many 0′s occur at the right of 1000!?

1.94 Prove that $(kn)!$ is divisible by $(n!)^k$ for all positive integers k and n.

1.95 Prove that $\binom{n}{k}$ is divisible by n if $\gcd(n, k) = 1$.

1.96 Let k and m be integers such that $0 \le m \le 2^k - 1$. Prove that the binomial coefficient $\binom{2^k - 1}{m}$ is odd.

1.97 Suppose that n has k 1′s when expressed in binary. Prove that the number of odd entries in the nth row of Pascal's dangle is 2^k.

1.98 Paul Erdős proved that there is only one binomial coefficient $\binom{n}{k}$ with $3 \le k \le n/2$ that is a power of an integer. Use a computer to find this binomial coefficient.

1.99 Use a computer to find all ordered pairs (n, e) with $2 \le e < (n - 1)/2 \le 50$ and $\Sigma_{k=0}^{e} \binom{n}{k}$ a power of 2. This calculation will be important in Chapters 5 and 6.

1.100 For what n is $n! + 1$ a perfect square?

1.101 Prove that $n!$ cannot be a perfect square greater than 1.

1.102 Notice that $6! = 3!5!$. Can you find other instances of integers a, b, and c, all greater than 1, such that $a!b! = c!$? Is there a pattern to these numbers?

1.103 Prove the following result of Erdős and Szekeres (1978):

$$\gcd\left(\binom{n}{i}, \binom{n}{j}\right) \ge 2^i,$$

where $0 < i \le j < n/2$.

1.104 (Fermat's little theorem) Prove that if p is a prime, then $a^P \equiv a \pmod{p}$. However, find a composite number n such that $2^n \equiv 2 \pmod{n}$.

1.105 Prove that $L_P \equiv 1 \pmod{p}$ if p is a prime. However, find a composite number n such that $L_n \equiv 1 \pmod{n}$.

1.106 (Perrin's sequence) Define $\{a_n\}$ by $a_0 = 3$, $a_1 = 0$, $a_2 = 2$, and $a_n = a_{n-3} + a_{n-2}$, for $n \geq 3$. This sequence is called *Perrin's sequence*.

Prove that $p|a_p$, for p prime. However, find a composite number n such that $n|a_n$.

1.107 Prove that $F_m|F_n$ if and only if $m|n$.

1.108 Prove that $\gcd(F_m, F_n) = F_{\gcd}(m,n)$.

Notes

Pascal's triangle is named after the French mathematician Blaise Pascal (1623–1662), who introduced it in his work *Trait du triangle arithmétique* (*Treatise on arithmetical triangle*), in which he used the triangle to solve problems in probability. However, the triangle was known much earlier in various places around the world. The Indian mathematician Bhattotpala (c. 1068) gave the first sixteen rows of the triangle. In China, the triangle is called "Yanghui's triangle," named after the mathematician Yang Hui (1238–1298).

The inclusion–exclusion principle was first studied by D. A. da Silva in 1854. It was studied by James Sylvester (1814–1897) in 1883 and is sometimes referred to as Sylvester's cross-classification principle.

Fibonacci (Leonardo of Pisa) introduced and discussed the Fibonacci numbers in his *Liber Abaci* ("Book of Calculations") in 1202. Abraham de Moivre gave the explicit formula for the Fibonacci numbers in 1730.

Lucas numbers were first investigated by François Édouard Anatole Lucas (1842–1891).

CHAPTER 2

GENERATING FUNCTIONS

Generating functions are algebraic objects that provide a powerful tool for analyzing recurrence relations. In this chapter, we will cover the basic theory of generating functions and examine many specific examples.

2.1 Rational generating functions

Given any sequence $a_0, a_1, a_2, \ldots$, the *ordinary generating function* is

$$f(x) = \sum_{n=0}^{\infty} a_n x^n. \tag{2.1}$$

The generating function $f(x)$ contains all of the information about the sequence $\{a_n\}$, and, being an algebraic entity, it is often easier to manipulate than the sequence itself. The term a_n is recovered by finding the coefficient of x^n in $f(x)$.

EXAMPLE 2.1 Generating function for the Fibonacci sequence

Find the ordinary generating function for the Fibonacci sequence $\{F_0, F_1, F_2, \ldots\}$.

Solution: Let $f(x) = \sum_{n=0}^{\infty} F_n x^n$. We obtain

$$\begin{aligned} f(x) &= x + x^2 + 2x^3 + 3x^4 + 5x^5 + \cdots \\ xf(x) &= x^2 + x^3 + 2x^4 + 3x^5 + 5x^6 + \cdots \\ x^2 f(x) &= x^3 + x^4 + 2x^5 + 3x^6 + 5x^7 + \cdots . \end{aligned}$$

Through mass cancellation, the recurrence relation for the Fibonacci numbers yields

$$f(x) - xf(x) - x^2 f(x) = x$$

and

$$f(x) = \frac{x}{1 - x - x^2}. \tag{2.2}$$

The function $f(x)$ contains complete information about the Fibonacci numbers and can be used to evaluate related infinite sums such as $\sum_{n=1}^{\infty} nF_n/3^n$. The computation of this sum is called for in the exercises.

For what values of x is the generating function valid?

Similarly, you can show that the generating function for the sequence of Lucas numbers $\{L_n\}$ is

$$\frac{2 - x}{1 - x - x^2}.$$

Notice that the generating function for the Fibonacci sequence is a rational function. Recall that a sequence $\{a_n\}$ satisfies a *linear recurrence relation with constant coefficients* $c_1,\ldots,c_k$ if

$$a_n = \sum_{i=1}^{k} c_i a_{n-i}$$

for all $n \geq k$. A sequence satisfies a linear recurrence relation with constant coefficients if and only if it has a rational ordinary generating function.

Also, notice that the denominator of the generating function for the Fibonacci numbers, $1 - x - x^2$, takes its form from the Fibonacci recurrence, while the numerator comes from multiplying the generating function by the denominator and keeping only those terms of degree less than 2.

Linear recurrence relation theorem. Let $\{a_n\}$ be a sequence and $c_1,\ldots,c_k$ be arbitrary numbers. Then the following three assertions are equivalent.

(1) The sequence $\{a_n\}$ satisfies a linear recurrence relation with constant coefficients c_1, $\ldots,c_k$, i.e.,

$$a_n = \sum_{i=1}^{k} c_i a_{n-i}, \quad n \geq k.$$

(2) The sequence $\{a_n\}$ has a rational ordinary generating function of the form

$$\frac{g(x)}{1 - \sum_{i=1}^{k} c_i x^i},$$

where g is a polynomial of degree at most $k - 1$.

(3) If

$$1 - \sum_{i=1}^{k} c_i x^i = (1 - r_1 x)(1 - r_2 x) \cdots (1 - r_k x),$$

with then r_i distinct, then the terms of $\{a_n\}$ are given by the formula

$$a_n = \alpha_1 r_1^n + \cdots + \alpha_k r_k^n, \quad n \geq 0,$$

where $\alpha_1,\ldots,\alpha_k$ are constants.

More generally, if

$$1 - \sum_{i=1}^{k} c_i x^i = (1 - r_1 x)^{m_1} (1 - r_2 x)^{m_2} \ldots (1 - r_l x)^{m_l},$$

where the roots $r_1,\ldots,r_1$ occur with multiplicities $m_1,\ldots,m_1$ then

$$a_n = p_1(n) r_1^n + \cdots + p_l(n) r_l^n, \quad n \geq 0,$$

where the p_i are polynomials with $\deg p_i < m_i$ for $1 \leq i \leq l$.

Proof. (1) $\Rightarrow$ (2) Assume that (a_n) satisfies a recurrence relation of the type specified in (1). Let $f(x)$ be the ordinary generating function for (a_n). Then

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} a_n x^n \\ &= \sum_{n=0}^{k-1} a_n x^n + \sum_{n=k}^{\infty} a_n x^n \\ &= \sum_{n=0}^{k-1} a_n x^n + \sum_{n=k}^{\infty} \sum_{i=1}^{k} c_i a_{n-i} x^n \\ &= \sum_{n=0}^{k-1} a_n x^n + \sum_{i=1}^{k} c_i \sum_{n=k}^{\infty} a_{n-i} x^n \\ &= \sum_{n=0}^{k-1} a_n x^n + \sum_{i=1}^{k} c_i x^i \sum_{n=k-i}^{\infty} a_n x^n \\ &= \sum_{n=0}^{k-1} a_n x^n + \sum_{i=1}^{k} c_i x^i \left(f(x) - \sum_{n=0}^{k-i-1} a_n x^n \right). \end{aligned}$$

Hence

$$f(x)\left(1 - \sum_{i=1}^{k} c_i x^i\right) = \sum_{n=0}^{k-1} a_n x^n - \sum_{i=1}^{k} c_i x^i \sum_{n=0}^{k-i-1} a_n x^n,$$

so that

$$f(x) = \frac{g(x)}{1 - \sum_{i=1}^{k} c_i x^i},$$

where $\deg g \le k - 1$.

(2) ⇒ (3) First consider the case where

$$1 - \sum_{i=1}^{k} c_i x^i = (1 - r_1 x) \cdots (1 - r_k x)$$

and then r_i are distinct. Expanding by partial fractions, we obtain

$$\begin{aligned} f(x) &= \frac{\alpha_1}{1 - r_1 x} + \cdots + \frac{\alpha_k}{1 - r_k x} \\ &= \alpha_1 \sum_{n=0}^{\infty} r_1^n x^n + \cdots + \alpha_k \sum_{n=0}^{\infty} r_k^n x^n \\ &= \sum_{n=0}^{\infty} (\alpha_1 r_1^n + \cdots + \alpha_k r_k^n) x^n, \end{aligned}$$

where $\alpha_1,\ldots,\alpha_k$ are constants. Since this is just another formula for the ordinary generating function for $\{a_n\}$, we have

$$a_n = \alpha_1 r_1^n + \cdots + \alpha_k r_k^n, \quad n \ge 0.$$

More generally, assume that $1 - \sum_{i=1}^{k} c_i x^i$ has repeated roots. Suppose that r is a root with multiplicity m. Then, in the partial fraction decomposition of

$$\frac{g(x)}{1-\sum_{i=1}^{k} c_i x^i},$$

we have terms

$$\frac{\beta_1}{(1-rx)}, \frac{\beta_2}{(1-rx)^2}, \ldots, \frac{\beta_m}{(1-rx)^m},$$

where $\beta_1, \beta_2, \ldots, \beta_m$ are constants. By the formula

$$\frac{1}{(1-x)^d} = \sum_{k=0}^{\infty} \binom{d+k-1}{k} x^k,$$

the contribution to x^i in the power series for these fractions is $p(n)x^n$, where p is a polynomial of degree less than m. The rest of the proof that (2) ⇒ (3) follows as in the special case.

Each step in the proof is reversible.

The factorization of $1 - \sum_{i=1}^{k} c_i x^i$ called for in the proof (and in practice) can be accomplished using the change of variables $y = 1/x$. Then

$$1 - \sum_{i=1}^{k} c_i x^i = 1 - \sum_{i=1}^{k} c_i y^{-i} = y^{-k}\left(y^k - \sum_{i=1}^{k} c_i y^{k-i}\right).$$

The problem is reduced to factoring the polynomial

$$y^k - \sum_{i=1}^{k} c_i y^{k-i}.$$

This polynomial is the characteristic polynomial of the recurrence relation.

EXAMPLE 2.2

Find the generating function for the sequence defined by the recurrence relation $a_n = 6a_{n-1} - 9a_{n-2}$ for $n \geq 2$, and $a_0 = 1$, $a_1 = 1$. (This comes from Example 1.31.) Use the generating function to find a direct formula for a_n.

Solution: The form of the recurrence relation tells us that the denominator of the generating function is $1 - 6x + 9x^2$. To get the numerator, we calculate

$$(1 - 6x + 9x^2)(a_0 + a_1 x) = (1 - 6x + 9x^2)(1 + x) = 1 - 5x + \cdots.$$

The only terms of degree less than 2 are $1 - 5x$, so the numerator is $1 - 5x$. Hence, the generating function is

$$\frac{1-5x}{1-6x+9x^2}.$$

To find a direct formula for a_n, we write the generating function as

$$(1-5x)(1-3x)^{-2}.$$

Thus, we have a binomial series with a negative exponent. Its expansion is

$$(1-5x)\sum_{k=0}^{\infty}(-1)^k 3^k \binom{-2}{k} x^k.$$

Therefore

$$a_n = (-1)^n 3^n \binom{-2}{n} - 5(-1)^{n-1} 3^{n-1} \binom{-2}{n-1}$$
$$= 3^n \binom{n+1}{n} - 5 \cdot 3^{n-1} \binom{n}{n-1}$$
$$= 3^n (n+1) - 5n3^{n-1}$$
$$= 3^n - 2n3^{n-1}, \quad n \geq 0.$$

This is the same solution we saw before.

EXAMPLE 2.3 Change for a dollar

How many ways can you make change for $1.00 using units of 0.01, 0.05, 0.10, 0.25, 0.50, and 1.00? Here are some examples:

$5 + 10 + 10 + 25 + 50$

$1 + 1 + 1 + 1 + 1 + 5 + 10 + 10 + 10 + 10 + 25 + 25$

$25 + 25 + 25 + 25.$

Solution: For $n \geq 0$, let a_n be the number of ways to make change for an amount n. We set $a_0 = 1$. The generating function for $\{a_n\}$ is

$$f(x) = 1 + 1x + 1x^2 + 1x^3 + 1x^4 + 2x^5 + 2x^6 + 2x^7 + 2x^8 + 2x^9 + 4x^{10} + \cdots$$

This generating function has the rational form

$$f(x) = \frac{1}{(1-x)(1-x^5)(1-x^{10})(1-x^{25})(1-x^{50})(1-x^{100})}. \tag{2.3}$$

Using a computer algebra system, we find that the coefficient of x^{100} of the generating function is 293, i.e., there are 293 ways to make change for a dollar.

The factors in the denominator of the generating function give rise to geometric series. For example, the second factor gives

$$\frac{1}{1-x^5} = 1 + x^5 + x^{2\cdot 5} + x^{3\cdot 5} + x^{4\cdot 5} + x^{5\cdot 5} + x^{6\cdot 5} + \cdots.$$

Each term in the product corresponds to a way to make change for a dollar. For example, the term corresponding to the sum 5 + 10 + 10 + 25 + 25 + 25 is shown in boldface:

$$(\mathbf{1} + x + x^2 + x^3 + x^4 + x^5 + x^6 + \cdots)$$
$$\cdot (1 + \mathbf{x^5} + x^{2\cdot 5} + x^{3\cdot 5} + x^{4\cdot 5} + x^{5\cdot 5} + x^{6\cdot 5} + \cdots)$$
$$\cdot (1 + x^{10} + \mathbf{x^{2\cdot 10}} + x^{3\cdot 10} + x^{4\cdot 10} + x^{5\cdot 10} + x^{6\cdot 10} + \cdots)$$
$$\cdot (1 + x^{25} + x^{2\cdot 25} + \mathbf{x^{3\cdot 25}} + x^{4\cdot 25} + x^{5\cdot 25} + x^{6\cdot 25} + \cdots)$$
$$\cdot (\mathbf{1} + x^{50} + x^{2\cdot 50} + x^{3\cdot 50} + x^{4\cdot 50} + x^{5\cdot 50} + x^{6\cdot 50} + \cdots)$$
$$\cdot (\mathbf{1} + x^{100} + x^{2\cdot 100} + x^{3\cdot 100} + x^{4\cdot 100} + x^{5\cdot 100} + x^{6\cdot 100} + \cdots).$$

Since the denominator of the generating function is a polynomial of degree 191, the sequence $\{a_n\}$ satisfies a linear recurrence relation of order 191. The explicit formula for a_n involves a sum of exponential functions which are powers of 100th roots of unity.

EXAMPLE 2.4 Alcuin's sequence

How many incongruent triangles have integer side lengths and perimeter 10^{100}?

Solution: Let $t(n)$ be the number of triangles with integer side lengths and perimeter n.

We set $t(0) = 0$. It is easy to find the values in the following table:

n	0	1	2	3	4	5	6	7	8
$t(n)$	0	0	0	1	0	1	1	2	1

The sequence $\{t(n)\}$ is known as Alcuin's sequence, named after Alcuin of York (735–804), who wrote a problem solving book containing some allocation problems equivalent to finding integer triangles.

The generating function for $\{t(n)\}$ is rational:

$$\sum_{n=0}^{\infty} t(n)x^n = \frac{x^3}{(1-x^2)(1-x^3)(1-x^4)}. \tag{2.4}$$

We prove this by showing that we can obtain any integer triangle from (1,1,1) by adding nonnegative integer multiples of (0,1,1), (1,1,1), and (1,1,2). These triples satisfy the weak triangle inequality, which is sufficient since we start with a genuine triangle. The unique solution to the equation

$$(a,b,c) = (1,1,1) + \alpha(0,1,1) + \beta(1,1,1) + \gamma(1,1,2),$$

in nonnegative integers α, β, and γ is $\alpha = b - a$, $\beta = a + b - c - 1$, $\gamma = c - b$. Since $2\alpha + 3\beta + 4\gamma = a + b + c - 3 = n - 3$, we see that $t(n)$ is equal to the number of ways of writing $n - 3$ as a sum of 2′s, 3′s, and 4′s (order of terms is unimportant). This is equivalent to the number of ways to make change for $n - 3$ using any combination of 2′s, 3′s, and 4′s.

Since the denominator of the (rational) generating function is

$$(1-x^2)(1-x^3)(1-x^4) = 1 - x^2 - x^3 - x^4 + x^5 + x^6 + x^7 - x^9,$$

the sequence $\{t(n)\}$ satisfies the order nine linear recurrence relation

$$t(n) = t(n-2)+t(n-3)+t(n-4)-t(n-5)-t(n-6)-t(n-7)+t(n-9), \quad n \geq 9.$$

The initial values, for $0 \leq n \leq 8$, are given in our table. The recurrence relation allows us to compute moderately large values of $t(n)$ easily by computer.

The generating function has the partial fraction decomposition

$$-\frac{1}{24(x-1)^3} + \frac{13}{288(x-1)} - \frac{1}{16(x+1)^2} - \frac{1}{32(x+1)} - \frac{x+1}{8(x^2+1)} + \frac{x+2}{9(x^2+x+1)}.$$

From this expansion we will find a simple formula for $t(n)$.

By the binomial series theorem, the first four terms can be written as

$$\frac{1}{24}\sum_{n=0}^{\infty}\binom{-3}{n}(-1)^n x^n - \frac{13}{288}\sum_{n=0}^{\infty}x^n - \frac{1}{16}\sum_{n=0}^{\infty}\binom{-2}{n}x^n - \frac{1}{32}\sum_{n=0}^{\infty}(-1)^n x^n.$$

From the identity $\binom{-k}{n} = (-1)^n \binom{n+k-1}{n}$, the coefficient of x^n is

$$\frac{1}{24}\binom{n+2}{n} - \frac{13}{288} - \frac{1}{16}(-1)^n(n+1) - \frac{1}{32}(-1)^n$$

$$= \frac{6n^2 + 18n - 1 - 18n(-1)^n - 27(-1)^n}{288}.$$

The final two terms yield coefficients of x^n that follow a pattern modulo 12, i.e., $c/72$ where c is given in the table:

$n \bmod 12$	0	1	2	3	4	5	6	7	8	9	10	11
c	7	−17	1	25	−17	−17	25	1	−17	7	1	1

Thus, we obtain the formula $t(n) = n^2/48 + (c - 7)/72$ for n even, and $t(n) = (n + 3)^2/48 + (c - 7)/72$ for n odd. This can be represented as

$$t(n) = \begin{cases} \left\{\frac{n^2}{48}\right\} & \text{for } n \text{ even} \\ \left\{\frac{(n+3)^2}{48}\right\} & \text{for } n \text{ odd,} \end{cases} \tag{2.5}$$

where $\{x\}$ is the nearest integer to x. For example, to find the number of integer triangles with perimeter 10^{100}, we compute $\{10^{200}/48\}$ by long division, and obtain the answer 2083 … 3, where there are one hundred ninety-six 3′s.

EXAMPLE 2.5

We can prove Lucas' theorem (see Section 1.10) using generating functions. Let's examine the case $k = 2$.

From the binomial theorem modulo p, we have

$$\begin{aligned} (1+x)^{a_0+a_1p+a_2p^2} &\equiv (1+x)^{a_0}(1+x)^{a_1p}(1+x)^{a_2p^2} \pmod p \\ &\equiv \left(1 + \binom{a_0}{1}x + \binom{a_0}{2}x^2 + \binom{a_0}{3}x^3 + \cdots\right) \\ &\quad \cdot \left(1 + \binom{a_1}{1}x^p + \binom{a_1}{2}x^{2p} + \binom{a_1}{3}x^{3p} + \cdots\right) \\ &\quad \cdot \left(1 + \binom{a_2}{1}x^{p^2} + \binom{a_2}{2}x^{2p^2} + \binom{a_2}{3}x^{3p^2} + \cdots\right) \pmod p \\ &\equiv \sum \binom{a_0}{b_0}\binom{a_1}{b_1}\binom{a_2}{b_2}x^{b_0+b_1p+b_2p^2} \pmod p. \end{aligned}$$

From the uniqueness of base b expansion, we conclude that

$$\binom{a_0}{b_0}\binom{a_1}{b_1}\binom{a_2}{b_2} \equiv \binom{a_0 + a_1p + a_2p^2}{b_0 + b_1p + b_2p^2} \pmod p.$$

EXERCISES

2.1 Evaluate the infinite series $\Sigma^{\infty}_{n=1} nF_n/3^n$.

2.2 Find a rational generating function for the sequence $\{a_n\}$ given by the recurrence formula

$$a_0 = 0,\ a_1 = 1,\ a_n = 5a_{n-1} - 6a_{n-2},\ n \geq 2.$$

2.3 Evaluate the infinite series

$$\sum_{n=0}^{\infty}(-1)^n \frac{a_n}{10^n},$$

where $\{a_n\}$ is the sequence of the previous exercise.

2.4 Find a rational generating function for the sequence of perfect squares, $\{n^2\}$, for $n \geq 0$.

2.5 Find a rational generating function for the sequence $\{a_n\}$ defined by $a_0 = 1$, $a_1 = 3$, $a_n = a_{n-1} + a_{n-2} + 2^{n-2}$, for $n \geq 2$. See Exercise 1.75.

2.6 (a) Suppose that $f(x)$ is the generating function for a sequence $\{a_n\}$. Show that

$$a_n = \frac{f^{(n)}(0)}{n!}.$$

(b) Suppose that $f(x, y)$ is the generating function (in two variables) for a sequence $\{a_{m,n}\}$. Show that

$$a_{m,n} = \frac{\partial_x^m \partial_y^n f(0,0)}{m!n!}.$$

2.7 Use a computer and an appropriate generating function to determine the number of ways of making change for \$1 using an even number of coins.

2.8 Use a generating function to determine the number of solutions in nonnegative integers to the equation

$$a + 2b + 4c = 10^{30}.$$

This problem duplicates Exercise 1.53.

2.9 Determine the number of solutions in nonnegative integers to the equation

$$a + 2b + 3c = 10^{30}.$$

2.10 Determine the number of solutions in nonnegative integers to the equation

$$a + b + 4c = 10^{30}.$$

2.11 (a) Show that the generating function (in two variables) for binomial coefficients is

$$\frac{1}{1 - x - y}.$$

(b) Show that the generating function (in three variables) for multinomial coefficients of the form (k_1,k_2,k_3^n) is

$$\frac{1}{1 - x - y - z}.$$

2.12 Find a recurrence formula for the coefficient of x^n in the series expansion of $(1 + 3x + x^2)^{-1}$.

2.13 Find a recurrence formula for the coefficient of x^n in the series expansion of $(1 + x + 2x^2)^{-1}$.

2.14 (Series multisection) Let $f(x) = \Sigma^{\infty}_{n=0} a_n x^n$. Show that, for any positive integers p and q with $0 \leq p < q$, we have

$$\sum_{k=0}^{\infty} a_{qk+p} x^{qk+p} = \frac{1}{q} \sum_{j=0}^{q-1} \omega^{-jp} f(\omega^j x),$$

where ω is a primitive qth root of unity.

2.15 Prove that, for each positive integer k, there exists a monk polynomial $p(n)$ of degree $k + 1$ with integer coefficients such that

$$\sum_{i=1}^{n} i^k \binom{n}{i} = 2^{n-k} p(n).$$

2.16 Prove that Alcuin's sequence is a zigzag sequence (its values alternately rise and fall) for $n \geq 6$.

2.17 Prove that Alcuin's sequence $\{t(n)\}$ satisfies the recurrence relation

$$t(n) = 3t(n-12) - 3t(n-24) + t(n-36), \quad n \geq 36.$$

2.18 Let t_n be the number of triples (a, b, c), where a, b, c are nonnegative integers satisfying $a \leq b \leq c$, $a + b \geq c$, and $a + b + c = n$. Find the generating function for $\{t_n\}$ and use the generating function to find t_n for $0 \leq n \leq 6$.

2.19 Given any integer $m > 1$, prove that Alcuin's sequence $\{t(n)\}$ is periodic modulo m with period $12m$.

2.2 Special generating functions

In this section we investigate certain interesting sequences and their generating functions.

Given any sequence $a_0, a_1, a_2, \ldots$, we define the *exponential generating function*

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n x^n}{n!}. \tag{2.6}$$

Let $d(x)$ be the exponential generating function for the sequence $\{d_n\}$, where d_n is the number of derangements of n elements:

$$d(x) = \sum_{n=0}^{\infty} d_n \frac{x^n}{n!}.$$

We set $d_0 = 1$. From the recurrence relation (1.31), it follows that

$$(1-x)d'(x) = xd(x),$$

for

$$\begin{aligned}
(1-x)d'(x) &= \sum_{n=1}^{\infty} d_n \frac{x^{n-1}}{(n-1)!} - \sum_{n=1}^{\infty} d_n \frac{x^n}{(n-1)!} \\
&= \sum_{n=0}^{\infty} d_{n+1} \frac{x^n}{n!} - \sum_{n=1}^{\infty} d_n \frac{x^n}{(n-1)!} \\
&= \sum_{n=0}^{\infty} \frac{d_{n+1} - nd_n}{n!} x^n \\
&= \sum_{n=0}^{\infty} \frac{nd_{n-1}}{n!} x^n \\
&= \sum_{n=1}^{\infty} \frac{d_{n-1}}{(n-1)!} x^n \\
&= xd(x).
\end{aligned}$$

Separating variables, we obtain

$$\int \frac{d'(x)}{d(x)}\,dx = \int \frac{x}{1-x}\,dx,$$

and hence

$$d(x) = C\frac{e^{-x}}{1-x}.$$

The value $d_0 = 1$ implies that $C = 1$. Therefore

$$d(x) = \frac{e^{-x}}{1-x}. \tag{2.7}$$

As we mentioned earlier, the coefficients of the generating function give us a formula for the general term of the underlying sequence. In this case, we obtain the explicit formula

$$d_n = \sum_{i=0}^{n} (-1)^i \frac{n!}{i!}. \tag{2.8}$$

We now consider a famous sequence of numbers called *Catalan numbers*, named after the mathematician Eugène Charles Catalan (1814–1894).

The *Catalan number* C_n is the number of sequences of n A's and n B's which have the property that, reading the sequence from left to right, at each symbol the number of As seen thus far is always greater than or equal to the number of B's seen thus far.

The five valid sequences for $n = 3$ are

AAABBB

AABABB

ABABAB

ABAABB

ABAABB.

Hence $C_3 = 5$.

With a little work, we find the following values of C_n:

n	1	2	3	4
C_n	1	2	5	14

We set $C_0 = 1$.

Let $f(x) = \sum_{n=0}^{\infty} C_n x^n$ be the ordinary generating function for the Catalan numbers. The strategy is to find an equation satisfied by $f(x)$, solve the equation, and read off the coefficients. The Catalan numbers satisfy the recurrence relation

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-1-k}, \quad n \geq 1. \tag{2.9}$$

(Every valid sequence can be written uniquely in the form As_1Bs_2, where s_1 is a valid sequence with k A's and B's, and s_2 is a valid sequence with $n - 1 - k$ A's and B's.) This recurrence relation implies

$$f(x) = 1 + xf(x)^2.$$

From the quadratic formula and the fact that $f(0) = 1$, it follows that

$$f(x) = \frac{1-\sqrt{1-4x}}{2x}. \tag{2.10}$$

From the binomial series for $\sqrt{1-4x}$, we obtain

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{2.11}$$

It is easy to show from (2.11) that

$$C_n = \frac{2(2n-1)}{n+1}C_{n-1}, \quad n \geq 1. \tag{2.12}$$

The Catalan numbers occur in many settings, such as the following:

- C_n is the number of lattice paths in the first quadrant of the plane which start at (0, 0), end at $(2n, 0)$, and proceed at each step by $(\Delta x, \Delta y)$ (+1, +1) or (+1, −1).
- C_{n-1} is the number of binary search trees on n vertices.
- C_{n-2} is the number of ways to parenthesize a product of n terms.
- C_{n-2} is the number of triangulations of a convex n-gon with $n - 3$ nonintersecting diagonals.

Next, we determine and work with generating functions for the Stirling numbers of the first and second kinds and for the Bell numbers.

We observe a pattern in the generating function for the Stirling numbers of the first kind. For instance.

$$\sum_{k=1}^{3} \begin{bmatrix} 3 \\ k \end{bmatrix} x^k = 2x + 3x^2 + 1x^3 = x(x+1)(x+2).$$

Define the *failing factorial* function $x_{(n)}$ as

$$x_{(n)} = x(x-1)(x-2)\cdots(x-n+1) \tag{2.13}$$

and the *rising factorial* function $x^{(n)}$ as

$$x^{(n)} = x(x+1)(x+2)\cdots(x+n-1). \tag{2.14}$$

We set $x^{(0)} = x_{(0)} = 1$.

The polynomials $x_{(n)}$ and $x^{(n)}$ are related by

$$\begin{aligned} (-x)_{(n)} &= -x(-x-1)\cdots(-x-n+1) \\ &= (-1)^n x(x+1)\cdots(x+n-1) \\ &= (-1)^n x^{(n)}. \end{aligned} \tag{2.15}$$

The following theorem states that $x^{(n)}$ is the ordinary generating function for the Stirling numbers of the first kind.

Generating function for the Stirling numbers of the first kind.

$$x^{(n)} = \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k. \tag{2.16}$$

Proof. The proof is by induction on n. The case $n = 1$ is trivial: $x^{(1)} = x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x^1$.

Assuming the result for n, it follows that

$$\begin{aligned} x^{(n+1)} &= x^{(n)}(x+n) \\ &= \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k (x+n) \\ &= \sum_{k=1}^{n+1} \left(\begin{bmatrix} n \\ k-1 \end{bmatrix} + n \begin{bmatrix} n \\ k \end{bmatrix} \right) x^k \\ &= \sum_{k=1}^{n+1} \begin{bmatrix} n+1 \\ k \end{bmatrix} x^k, \end{aligned}$$

which is the correct formula for the $n + 1$ case.

EXAMPLE 2.6 Expected number of cycles in a permutation

What is the expected number of cycles in a randomly chosen permutation of $\{1,2,3,\ldots,n\}$?

Solution: Differentiating the generating function for $\begin{bmatrix} n \\ k \end{bmatrix}$ with respect to x, we obtain

$$\begin{aligned} D_x x^{(n)} &= D_x \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k \\ D_x[x(x+1)\cdots(x+n-1)] &= \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} k x^{k-1}. \end{aligned}$$

Evaluating the second identity at $x = 1$, we obtain

$$n! \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} k.$$

Dividing by $n!$, we find that the expected number of cycles is

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},$$

an expression asymptotic to $\ln n$. For instance, in a permutation on 1000 elements, we expect to find about seven cycles.

The polynomial $x_{(n)}$ is the generating function for the *signed Stirling numbers of the first kind*, $(-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix}$.

Generating function for the signed Stirling numbers of the first kind.

$$x_{(n)} = \sum_{k=1}^{n} (-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix} x^k. \tag{2.17}$$

Proof. We have

$$\begin{aligned} x_{(n)} &= (-1)^n(-x)^{(n)} \\ &= (-1)^n \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} (-x)^k \\ &= \sum_{k=1}^{n} (-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix} x^k. \end{aligned}$$

To test this generating function, we compute $x_{(3)} = x(x-1)(x-2) = x^3 - 3x^2 + 2x$ and note that the coefficients 1, –3, 2 are the signed Stirling numbers of the first kind with $n = 3$.

Generating function for the Stirling numbers of the second kind.

$$x^n = \sum_{k=1}^{n} \begin{Bmatrix} n \\ k \end{Bmatrix} x_{(k)}. \tag{2.18}$$

Proof. We have

$$\begin{aligned} x^n &= \sum_{k=1}^{n} T(n,k) \binom{x}{k} \\ &= \sum_{k=1}^{n} \frac{T(n,k)}{k!} x_{(k)} \\ &= \sum_{k=1}^{n} \begin{Bmatrix} n \\ k \end{Bmatrix} x_{(k)}. \end{aligned}$$

The vector space of polynomials with real coefficients has as bases the two sets $\mathcal{B}_1 = \{x^n : n \geq 0\}$ and $\mathcal{B}_2 = \{x_{(n)} : n \geq 0\}$. Recall that we have set $\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1$ and $\begin{Bmatrix} n \\ 0 \end{Bmatrix} = \begin{bmatrix} n \\ 0 \end{bmatrix} = 0$ for $n \geq 1$. Then $S_1 = [(-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix}]$ is the change of basis matrix from $\mathcal{B}_2$ to $\mathcal{B}_1$, while $S_2 = [\begin{Bmatrix} n \\ k \end{Bmatrix}]$ is the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$. Therefore, the two matrices S_1 and S_2 are inverses, so that $S_1S_2 = S_2S_1 = I$, where I is the infinite-dimensional identity matrix. In summation form, this assertion is written as

$$\sum_{k=1}^{n} (-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix} \begin{Bmatrix} k \\ j \end{Bmatrix} = \sum_{k=1}^{n} (-1)^{j+k} \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{bmatrix} k \\ j \end{bmatrix} = \delta(n,j). \tag{2.19}$$

Recall that $\delta(n, j) = 1$ if $n = j$ and 0 if $n \neq j$. This identity leads (as per Exercise 1.59) to the following wonderful inversion formula.

Inversion formula for summations with Stirling numbers as weights. For any two real-valued functions f and g, we have

$$g(n) = \sum_{k=1}^{n} \begin{Bmatrix} n \\ k \end{Bmatrix} f(k)$$

if and only if

$$f(n) = \sum_{k=1}^{n} (-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix} g(k).$$

Proof Assume that $g(n) = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} f(k)$. Then

$$\begin{aligned}
\sum_{k=1}^{n} (-1)^{n+k} \left[{n \atop k} \right] g(k) &= \sum_{k=1}^{n} \sum_{j=1}^{k} (-1)^{n+k} \left[{n \atop k} \right] \left\{ {k \atop j} \right\} f(j) \\
&= \sum_{j=1}^{n} \sum_{k=j}^{n} (-1)^{n+k} \left[{n \atop k} \right] \left\{ {k \atop j} \right\} f(j) \\
&= \sum_{j=1}^{n} \delta(n, j) f(j) \\
&= f(n).
\end{aligned}$$

The reverse implication is proved similarly.

Let t_n be the number of transitive and reflexive relations on an arbitrary n-set X. It can be shown that t_n is the number of topologies on an n-set. Let p_n be the number of partial orders on X. The reader may wish to verify that $p_1 = 1$, $p_2 = 3$, $p_3 = 19$ and $t_1 = 1$, $t_2 = 4$, $t_3 = 29$. Although there are no known formulas for t_n and p_n, the two functions are related by our inversion formula.

Suppose $X = \{1,\ldots,n\}$ and R is a transitive and reflexive relation on X. We define a new relation R' on X as follows: $(a, b) \in R'$ if and only if $(a, b) \in R$ and $(b, a) \in R$. We claim that R' is an equivalence relation on X. Certainly R' is reflexive, as $(a, a) \in R$ implies $(a, a) \in R'$. If $(a, b) \in R'$, then $(b, a) \in R'$ (by definition), so R' is symmetric. If $(a, b) \in R'$ and $(b, c) \in R'$, then (a, b), (b, c), (b, a), $(c, b) \in R$, which implies that (a, c), $(c, a) \in R$ (because R is transitive); hence, $(a, c) \in R'$, so R' is transitive.

Therefore, R' is an equivalence relation with equivalence classes $[a] = \{b \in X : (a, b) \in R \text{ and } (b, a) \in R\}$. This means that in order to construct a transitive, reflexive relation on X, we must first partition X into equivalence classes. As X may be partitioned into k equivalence classes in $f(x) = \sum_{n=0}^{\infty} F_n x^n$ ways, the question is, how are the equivalence classes pieced together? Suppose $[a]$ and $[b]$ are two different equivalence classes under R' and $(a, b) \in R$. By transitivity, $(a, c) \in R$ for all $c \in [b]$. Again, by transitivity, $(d, c) \in R$ for all $d \in [a]$. To paraphrase: "Everything in $[a]$ is arrowed to everything in $[b]$." Therefore we can think of the equivalence classes as points that are either joined or not joined by an arrow. This defines a partial order on the set of equivalence classes. In other words, the k equivalence classes are partially ordered. Because this may be done in p_k ways, summing over all possible values of k, we obtain

$$t_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} p_k. \tag{2.20}$$

We test this equation by putting in the values $p_1 = 1, p_2 = 3, p_3 = 19, \left\{ {3 \atop 1} \right\} = 1, \left\{ {3 \atop 2} \right\} = 3, \left\{ {3 \atop 3} \right\} = 1$, and calculating $t_3 = 29$.

Our equation would provide a formula for t_n if only a formula for p_n were known (as we already have a formula for $\left\{ {n \atop k} \right\}$). However, using the inversion formula for Stirling numbers we can write p_n in terms of t_n:

$$p_n = \sum_{k=1}^{n} (-1)^{n+k} \left[{n \atop k} \right] t_k. \tag{2.21}$$

For example, putting in the values $t_1 = 1, t_2 = 4, t_3 = 29, (-1)^{3+1}\begin{bmatrix}3\\1\end{bmatrix} = 2, (-1)^{3+2}\begin{bmatrix}3\\2\end{bmatrix} = -3, (-1)^{3+3}\begin{bmatrix}3\\3\end{bmatrix} = 1$, we obtain $p_3 = 19$.

Although no explicit formulas are known for the general terms of the sequences $\{p_n\}$ and $\{t_n\}$, it is known that $p_n \sim t_n$ and $\log_2 p_n = n^2/4 + o(n^2)$. We let p_n^* and t_n^* be the unlabeled set versions of p_n and t_n. There are no known formulas for these sequences, although it is known that $p_n^* \sim p_n/n!$. One might think that $t_n^* = \Sigma^n_{k=1} p(n, k)p_k^*$, but this is false. Why?

Open problem. Find explicit formulas for p_n, p_n^*, t_n, and t_n^*.

Table 2.1 presents the first few terms of the four sequences. See [24] or the Online Encyclopedia of Integer Sequences.

Table 2.1 Some terms of four important sequences.

n	1	2	3	4	5	6	7	8
p_n	1	3	19	219	4231	130023	6129859	431723379
p_n^*	1	2	5	16	63	318	2045	16999
t_n	1	4	29	355	6942	209527	9535241	642779354
t_n^*	1	3	9	33	139	718	4535	35979

Now we give the exponential generating function for the Bell numbers $B(n)$.

Generating function for the Bell numbers. $\Sigma^\infty_{n=0} B(n)x^n/n! = e^{e^x-1}$.

Proof. We have

$$\begin{aligned}
\sum_{n=0}^{\infty} \frac{B(n)x^n}{n!} &= e^{-1} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{j=0}^{\infty} \frac{j^n}{j!} x^n \\
&= e^{-1} \sum_{j=0}^{\infty} \frac{1}{j!} \sum_{n=0}^{\infty} \frac{(jx)^n}{n!} \\
&= e^{-1} \sum_{j=0}^{\infty} \frac{e^{jx}}{j!} \\
&= e^{-1} e^{e^x} \\
&= e^{e^x-1}.
\end{aligned}$$

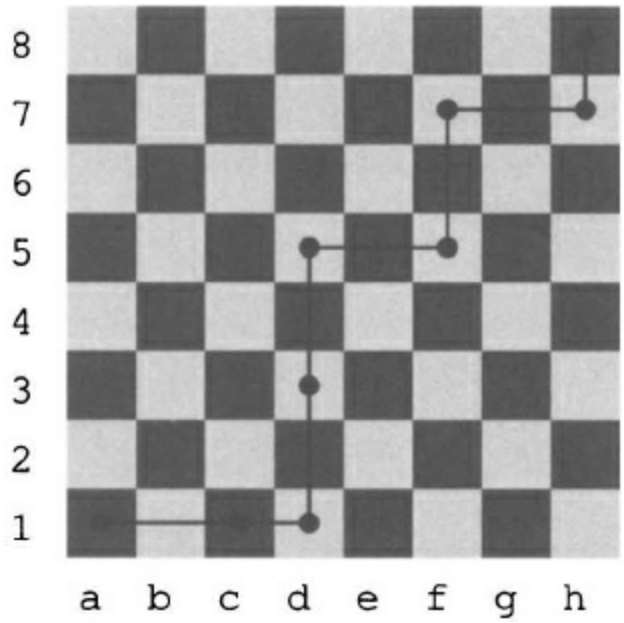
What an interesting-looking generating function! The reader may wish to compute the first four terms of the generating function and compare them to the known values $B(0) = 1$, $B(1) = 1$, $B(2) = 2$, $B(3) = 5$.

EXAMPLE 2.7 Rook walks

A chess Rook can move any number of squares horizontally or vertically on a chess board. How many different walks can a Rook travel in moving from the lower-left corner (al) to the upper-right corner (h8) on the board? Assume that the Rook moves right or up at every step. For

example, Figure 2.1 shows the Rook walk a1-c1-d1-d3-d5-f5-f7-h7-h8.

Figure 2.1 A Rook walk from a1 to h8.



Note that if the Rook could only move one square at a time, this problem would be equivalent to the problem of counting paths along city streets (see Example 1.14). Here the number of such paths is simply $\binom{14}{7} = 3432$. We expect that the number of Rook walks will be much larger.

Solution: We generalize the problem by considering Rook walks to any square on the board (with the Rook starting on a1 and moving toward the goal square at every step). We make a table displaying the number of walks. The bottom-left entry is the number of Rook walks from a1 to a1, which is 1. Each other entry is the sum of all the entries below or to the left of the given entry. The reason is that the Rook's last move must come from one of the squares represented by these entries. For example, the entry corresponding to the c4 square is 4 + 12 + 2 + 5 + 14 = 37. It's possible to complete the table by hand in a few minutes. The number of Rook walks from a1 to h8 is 470,010. (The dots in the table indicate that we can generalize the problem to arbitrarily large chess boards.)

⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
64	320	1328	4864	16428	52356	159645	470010	...
32	144	560	1944	6266	19149	56190	159645	...
16	64	232	760	2329	6802	19149	52356	...
8	28	94	289	838	2329	6266	16428	...
4	12	37	106	289	760	1944	4864	...
2	5	14	37	94	232	560	1328	...
1	2	5	12	28	64	144	320	...
1	1	2	4	8	16	32	64	...

Let $a(m, n)$ be the number of Rook walks from $(0, 0)$ to (m, n), and set $a(m, n) = 0$ if $m < 0$ or $n < 0$. The generating function for the doubly infinite sequence $\{a(m, n)\}$ is the rational function

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a(m,n)s^m t^n = \frac{1}{1-\frac{s}{1-s}-\frac{t}{1-t}} = \frac{1-s-t+st}{1-2s-2t+3st}. \tag{2.22}$$

Can you explain why?

It follows that the sequence $a(m, n)$ satisfies a recurrence relation (with initial values):

$$a(m,n) = 2a(m,n-1) + 2a(m-1,n) - 3a(m-1,n-1), \quad m \geq 2 \text{ or } n \geq 2;$$
$$a(0,0) = 1,\ a(0,1) = 1,\ a(1,0) = 1,\ a(1,1) = 2.$$

Can you explain why this recurrence relation holds by counting Rook walks?

Let $a_n = a(n, n)$, for $n \geq 0$. The sequence $\{a_n\}$, called the diagonal sequence, is

1, 2, 14, 106, 838, 6802, 56190, 470010, ...

The eighth term, 470,010, is the number of Rook walks in the original problem.

Let $f(x)$ be the generating function for the diagonal sequence, i.e.,

$$f(x) = 1 + 2x + 14x^2 + 106x^3 + 838x^4 + 6802x^5 + 56190x^6 + 470010x^7 + \cdots.$$

We will show that

$$f(x) = \frac{1}{2}\left(1 + \frac{\sqrt{1-x}}{\sqrt{1-9x}}\right). \tag{2.23}$$

In order to get the generating function for the diagonal sequence, we make the change of variables $t = x/s$. To include terms such as s^3t^5, we allow arbitrary integer exponents for s. For instance, we represent s^3t^5 as $s^{-2}x^5$. The diagonal generating function is the coefficient of s^0. For more about this method, see [25].

We obtain the generating function

$$\frac{1 - s - x/s + x}{1 - 2s - 2x/s + 3x} = \frac{1}{2}\left(1 + \frac{(1-x)s}{-2s^2 + (3x+1)s - 2x}\right).$$

We now consider the function

$$\frac{s}{-2s^2 + (3x+1)s - 2x}.$$

Using the quadratic formula, we write this as

$$\frac{s}{-2(s-\alpha)(s-\beta)},$$

where

$$\alpha = \frac{3x + 1 - \sqrt{(1-x)(1-9x)}}{4},\ \beta = \frac{3x + 1 + \sqrt{(1-x)(1-9x)}}{4}.$$

We put our function into partial fraction form,

$$\frac{1}{2(\beta-\alpha)}\left[\frac{\alpha}{s-\alpha} - \frac{\beta}{s-\beta}\right],$$

or

$$\frac{1}{2(\beta-\alpha)}\left[\frac{\alpha/s}{1-(\alpha/s)} + \frac{1}{1-(s/\beta)}\right].$$

For $-1/9 < x < 1/9$, we expand the function as a Laurent series in the annulus $|\alpha| < |s| < |\beta|$ in powers of α/s and s/β, obtaining

$$\frac{1}{2(\beta-\alpha)}\left[\sum_{n=1}^{\infty}\left(\frac{\alpha}{s}\right)^n + \sum_{n=0}^{\infty}\left(\frac{s}{\beta}\right)^n\right].$$

The coefficient of s^0 in this series is

$$\frac{1}{2(\beta-\alpha)} = \frac{1}{\sqrt{(1-x)(1-9x)}}.$$

This establishes the generating function for the diagonal sequence.

We can use the generating function for the diagonal sequence to obtain a recurrence relation for it. By inspection,

$$f'(x)(1-x)(1-9x) = 4f(x) - 2.$$

We can read off a recurrence relation for $\{a_n\}$ by looking at the coefficient of x^n in the above relation:

$$a_0 = 1,\ a_1 = 2$$

$$a_n = \frac{(10n-6)a_{n-1} - (9n-18)a_{n-2}}{n}, \quad n \geq 2. \tag{2.24}$$

A counting argument for this relation is provided by E. Y. Jin and M. E. Nebel in "A combinatorial proof of the recurrence for rook paths," in *Electronic Journal of Combinatorics*, **19**, no. 1 (2012).

We can also use the generating function to determine the asymptotic growth rate of a_n. We have

$$a_n \sim \frac{\sqrt{2}}{3} \cdot \frac{9^n}{\sqrt{\pi n}}. \tag{2.25}$$

See Exercises.

Can you find the generating function for the number of King walks from the lower-left corner to the upper-right corner of an arbitrary-size chess board? At each step, the King moves one square to the right, one square up, or both. Such walks are called *Delannoy walks*.

EXERCISES

2.20 Prove that

$$\sum_{k=0}^{[n/2]} \left[\binom{n}{k} - \binom{n}{k-1} \right]^2 = \frac{1}{n+1}\binom{2n}{n}.$$

2.21 Prove that the Catalan number C_n is odd if and only if $n = 2^k - 1$ for some positive integer k.

2.22 Prove that for any positive integer n we have

$$C_{3n-1} \equiv C_{3n} \equiv C_{3n+1} \pmod 3.$$

2.23 Prove that if $n > 3$, then the Catalan number C_n is not a prime.

2.24 Define the $n \times n$ matrix $A = [a_{ij}]$ by the rule $a_{ij} = C_{i+j-2}$. Prove that $\det A = 1$.

2.25 (Ising problem, $2 \times n$ case) Let $O(a, b)$ be a box consisting of $a + b$ cells, each 1×1, arranged in a row of length a sitting on top of a row of length b (the leftmost cells of the two rows line up). Let $f(a, b)$ be the number of ways of covering $O(a, b)$ with 1×1 and 1×2 tiles. Set up a recurrence relation for $f(a, b)$. Generate some data using a computer and make a conjecture about a formula for $f(a, b)$.

2.26 Prove that $\sum_{k=0}^{n} (-1)^{n+k} \begin{bmatrix} n \\ k \end{bmatrix} B(k) = 1$.

2.27 Prove that

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

and

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

2.28 Let X be the number of cycles of a randomly chosen permutation of $\{1,\ldots,n\}$. We have shown that $E(X) \sim \ln n$. Show that $\mathrm{Var}(X) \sim \ln n$.

Hint: From the relation $x^{(n)} = \sum_{k=1}^{n} \left[{n \atop k}\right] x^k$ obtain

$$\mathrm{Var}(X) = \left. \frac{1}{n!} \frac{d}{dx} x \frac{d}{dx} x^{(n)} \right]_{x=1}.$$

Use logarithmic differentiation.

2.29 Let a_n be the number of permutations in S_n which alternately rise, fall, rise, fall, etc. For example, 142635 is such a permutation. Find $\sum_{n=1}^{\infty} a_n x^n/n!$ and use this information to find a_6.

2.30 Let $\{a_n\}$ be a sequence with the exponential generating function

$$\sum_{n=0}^{\infty} \frac{a_n x^n}{n!} = \exp\left(x + \frac{x^2}{2}\right).$$

Evaluate the sum

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

Interpret the sum combinatorially. Using Exercise 1.58, find a combinatorial interpretation of a_n. See Exercises 1.37, 2.45, and 2.69.

Hint: Multiply both sides by e^{-x}.

2.31 A lone King is on a chessboard. How many paths may the King take from the lower-left corner (al) of the board to the lower-right corner (h8), moving one square right, up, or up-right at every step?

2.32 Find a recurrence relation for the number of Queen walks from the lower-left corner of an arbitrary-size chess board to the square (n, n). (A Queen can move any number of squares horizontally, vertically, or diagonally. Assume that the Queen always moves to the right, up, or up-right.) Find the corresponding generating function.

2.33 Suppose that a ChildRook can move like a chess Rook but only at most two squares horizontally or vertically. Let $a(m, n)$ be the number of walks a ChildRook can take from square (1, 1) to square (m, n) on an arbitrary-size chess board. (Assume that the RookPlus always moves toward the goal square.) Find a finite-order recurrence relation for $a(m, n)$.

2.34 Suppose that a RookPlus can move like a chess Rook with the additional possibility of moving one square diagonally. Let $b(m, n)$ be the number of walks a RookPlus can take from square (1, 1) to square (m, n) on an arbitrary-size chess board. (Assume that the RookPlus always moves toward the goal square.) Find a finite-order recurrence relation for $b(m, n)$.

2.35 Show that

$$a_n \sim \frac{\sqrt{2}}{3} \cdot \frac{9^n}{\sqrt{\pi n}},$$

where a_n is the number of Rook walks from (0, 0) to (n, n).

Hint: We know that a_n, for $n \geq 1$, is the coefficient of x^n in the generating function $\frac{1}{2}\sqrt{1-x}/\sqrt{1-9x}$. We examine this generating function, which has a singularity at $x = 1/9$. Let $y = 9x$ and $h(y) = \frac{1}{2}\sqrt{1-y/9} = h_0 + h_1 y + h_2 y^2 + \cdots$. Then $h_0 + h_1 + h_2 + \cdots = h(1) = \sqrt{2}/3$. Suppose that $(1-y)^{-1/2} = s_0 + s_1 x + s_2 x^2 + \cdots$. Then

$$a_n = (h_0 s_n + h_1 s_{n-1} + \cdots + h_n s_0) 9^n \sim h(1) s_n 9^n,$$

and, using Stirling's approximation $n! \sim n^n e^{-n}\sqrt{2\pi n}$, we have

$$s_n = \binom{-\frac{1}{2}}{n}(-1)^n \sim \frac{1}{\sqrt{\pi n}}.$$

2.3 Partition numbers

In Chapter 1, we defined the partition number $p(n, k)$ to be the number of ways to write n as a sum of k positive integers. We can also think of $p(n, k)$ as the number of onto functions $f: X \to Y$, $|X| = n$, $|Y| = k$, where X and Y are unlabeled sets. Also, the partition number $p(n)$ has been defined as $p(n) = \Sigma^n_{k=1} p(n, k)$. In this section we develop generating functions for partition numbers.

EXAMPLE 2.8

Determine $p(4, k)$, for $k = 1, 2, 3, 4$, and $p(4)$.

Solution: We have

$$p(4,1) = 1 \qquad (4 = 4)$$
$$p(4,2) = 2 \qquad (2+2=4,\ 3+1=4)$$
$$p(4,3) = 1 \qquad (2+1+1=4)$$
$$p(4,4) = 1 \qquad (1+1+1+1=4)$$

and

$$\begin{aligned} p(4) &= p(4,1) + p(4,2) + p(4,3) + p(4,4) \\ &= 1 + 2 + 1 + 1 \\ &= 5. \end{aligned}$$

Suppose that X and Y are unlabeled and $f: X \to Y$ is an onto function ($|X| = n$, $|Y| = k$). Suppose that the inverse images of the elements of Y have cardinalities $\lambda_1, \ldots, \lambda_k$. Because the inverse images account for all the elements of X, it follows that

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = n.$$

Furthermore, we may assume that the λ_i are ordered from largest to smallest:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0.$$

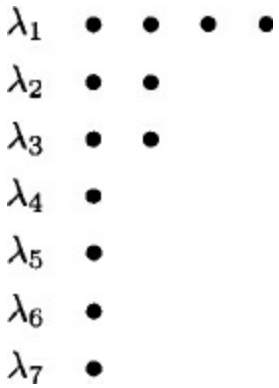
A partition $\lambda_1 + \cdots + \lambda_k = n$ may be pictured with a *Ferrers diagram* consisting of k rows of dots with λ_i dots in row i, where $1 \leq i \leq k$. The Ferrers diagram for the partition $7 + 3 + 1 + 1 =$

12 is shown in Figure 2.2.

Figure 2.2 The Ferrers diagram of a partition of 12.

The *transpose* of a Ferrers diagram is created by writing each row of dots as a column. The resulting partition is called the *conjugate* of the original partition. For example, the partition 12 = 7 + 3 + 1 + 1 of Figure 2.2 is transposed to create the conjugate partition 12 = 4 + 2 + 2 + 1 + 1 + 1 + 1 of Figure 2.3.

Figure 2.3 A transpose Ferrers diagram.



The reader may enjoy matching each partition of 4 on the previous page with its conjugate. (One partition is self-conjugate.)

We now give the ordinary generating function for the partition numbers $p(n)$. For convenience, we set $p(0) = 1$.

Generating function for integer partitions.

$$\sum_{n=0}^{\infty} p(n)x^n = \prod_{k=1}^{\infty} \frac{1}{1-x^k}.$$

Proof. We need to show that the coefficients of x^n on the two sides of the equation are equal. The coefficient of x^n on the left side is patently $p(n)$. On the right side, the product may be written as

$$\prod_{k=1}^{\infty} \frac{1}{1-x^k} = \prod_{k=1}^{\infty} (1 + x^k + x^{2k} + x^{3k} + x^{4k} + \cdots).$$

To find the contribution to x^n from this product, suppose the term $x^{m(k)k}$ is selected from the kth factor and these terms are multiplied to yield $x^{m(1)+m(2)2+\cdots}$. If this expression equals x^n, then

(2.26) $$m(1) + m(2)2 + \cdots = n.$$

Contributions to x^n correspond to solutions of (2.26). These solutions may be envisioned as Ferrers diagrams for partitions of n. With t as the greatest integer for which $m(t)$ is nonzero, we create the Ferrers diagram with $m(t)$ rows of t dots, followed by $m(t-1)$ rows of $t-1$ dots, etc. This correspondence between solutions of (2.26) and partitions of n completes the proof.

Determining the ordinary generating function for $p(n, k)$ with k fixed is not difficult. We start by

making an elementary observation from the transposes of Ferrers diagrams.

Theorem. The number of partitions of n into exactly k parts is equal to the number of partitions of n where the size of the greatest summand is k.

Now we can establish the generating function for $p(n, k)$.

Theorem.

$$\sum_{n=k}^{\infty} p(n,k)x^n = x^k \prod_{j=1}^{k} \frac{1}{1-x^j}.$$

Proof We define $p(n, \leq k)$ to be the number of partitions of n into at most k parts. From the previous theorem, $p(n, \leq k)$ is also the number of partitions of n into parts of size at most k. Clearly, $p(n, k) = p(n, \leq k) - p(n, \leq k-1)$. We obtain

$$\begin{aligned}\sum_{n=k}^{\infty} p(n,\leq k)x^n &= \prod_{j=1}^{k}(1+x^j+x^{2j}+x^{3j}+\cdots) \\ &= \prod_{j=1}^{k}\frac{1}{1-x^j}.\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{n=k}^{\infty} p(n,k)x^n &= \sum_{n=k}^{\infty}[p(n,\leq k)-p(n,\leq k-1)]x^n \\ &= \prod_{j=1}^{k}\frac{1}{1-x^j} - \prod_{j=1}^{k-1}\frac{1}{1-x^j} \\ &= \prod_{j=1}^{k}\frac{1}{1-x^j}[1-(1-x^k)] \\ &= x^k\prod_{j=1}^{k}\frac{1}{1-x^j}.\end{aligned}$$

Let $p(n, O)$ be the number of partitions of n into summands each of which is an odd number. Let $p(n, D)$ be the number of partitions of n into distinct summands. As a further illustration of the use of generating functions, we prove that $p(n, O) = p(n, D)$.

Euler's theorem (1748).

$$p(n,O) = p(n,D).$$

Proof. We have

$$\begin{aligned}\sum_{n=0}^{\infty} p(n,O)x^n &= \frac{1}{(1-x)(1-x^3)(1-x^5)\dots} \\ &= \frac{(1-x^2)}{(1-x)(1-x^2)} \cdot \frac{(1-x^4)}{(1-x^3)(1-x^4)} \cdot \frac{(1-x^6)}{(1-x^5)(1-x^6)} \cdots \\ &= \frac{(1-x^2)}{(1-x)} \cdot \frac{(1-x^4)}{(1-x^2)} \cdot \frac{(1-x^6)}{(1-x^3)} \cdots \\ &= (1+x)(1+x^2)(1+x^3)\dots \\ &= \sum_{n=0}^{\infty} p(n,D)x^n.\end{aligned}$$

The desired equality follows by comparing coefficients of the two generating functions.

EXERCISES

2.36 Show that the number of partitions of n into summands none of which occurs exactly once is the same as the number of partitions of n into summands none of which is congruent to 1 or 5 modulo 6.

2.37 Find formulas for $p(n, 1)$, $p(n, 2)$, and $p(n, 3)$. Find an asymptotic estimate for $p(n, k)$ (with k fixed).

2.38 Prove Jacobi's identity:

$$1 + \sum_{n=1}^{\infty} \frac{x^{n^2}}{(1-x)^2(1-x^2)^2\cdots(1-x^n)^2} = \prod_{k=1}^{\infty} \frac{1}{1-x^k}.$$

2.39 We say that two permutations σ and τ of S_n are in the same *conjugacy class* if there exists a permutation $\rho \in S_n$ such that $\tau = \rho\sigma\rho^{1}$. Prove that two permutations are in the same conjugacy class if and only if they have the same cycle structure. How many conjugacy classes of S_n are there?

2.40

(a) How many nonisomorphic abelian groups of order 2700 are there?

(b) How many ways may one make \$2.23 postage using 1 cent, 2 cent, 3 cent, 10 cent, 20 cent, \$1, and \$2 stamps, and not more than three of any one denomination?

Notice that the answers to parts (a) and (b) are the same. Why is this?

2.4 Labeled and unlabeled sets

Many enumeration problems can be solved by representing the objects to be counted as functions $f : X \to Y$, where X and Y are chosen appropriately. The conditions imposed in the enumeration problem usually amount to putting restrictions on the functions (e.g., requiring them to be one-to-one) and/or making rules as to when two functions are considered equivalent (e.g., when they are equivalent up to a permutation of the elements of X). For example, the binomial coefficient $\binom{n}{m}$ is the number of 1–1 functions from an m-set X into an n-set Y, where the elements of X are considered to be indistinct.

Suppose that X and Y are two finite nonempty sets and Y^X is the collection of functions $f : X \to Y$. We wish to define some equivalence relations on Y^X and, in each case, count the number of equivalence classes. When we say that two functions f and $g (f, g \in Y^X)$ are *equivalent*, we will mean one of four things:

(1) $f = g$.

(2) $f = gh$ for some bijection $h : X \to X$.

(3) $f = ig$ for some bijection $i : Y \to Y$.

(4) $f = igh$ for some bijections $h : X \to X$ and $i : Y \to Y$.

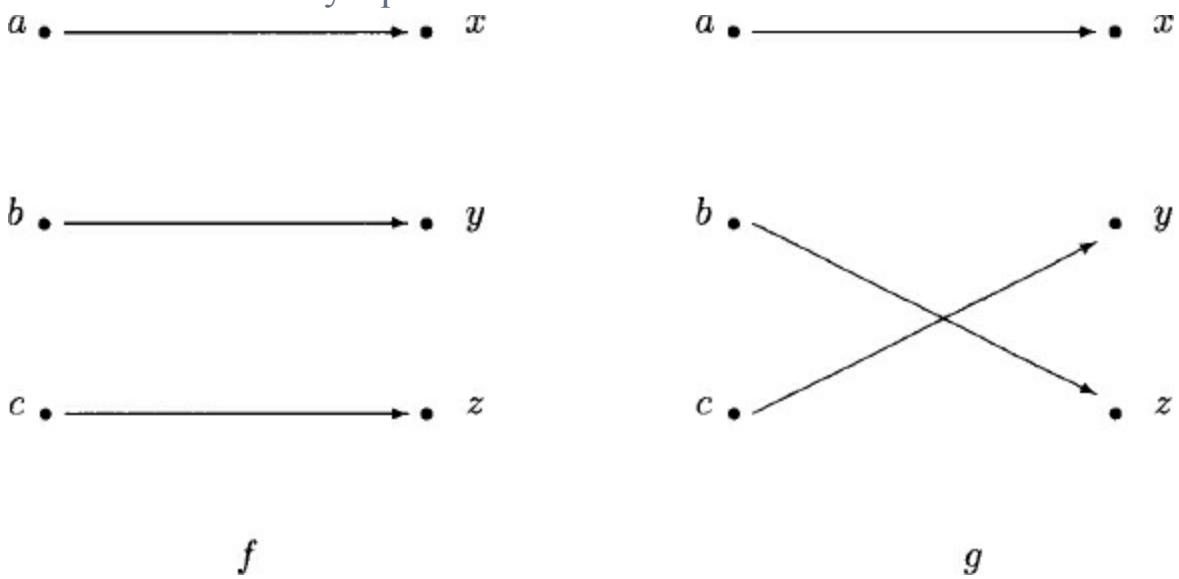
(Note that the functions here are applied from right to left.)

In definitions (2) and (4) we say that h *delabels* X, and we speak of X as an *unlabeled* (or *delabeled*) set. Likewise, in definitions (3) and (4) we say that i delabels Y, and we speak of Y as an unlabeled or delabeled set.

EXAMPLE 2.9

Consider the functions f and g of Figure 2.4. Are they equivalent?

Figure 2.4 Two functions. Are they equivalent?

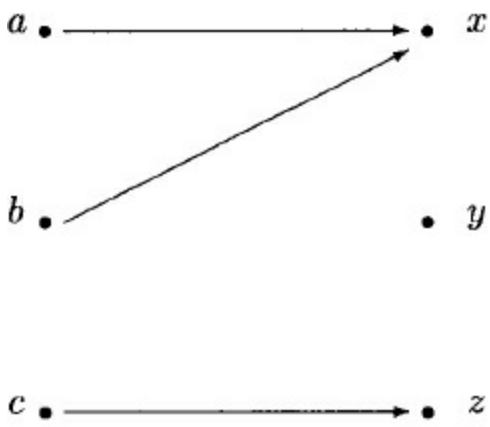


Solution: The functions fail to be equivalent under definition (1) because, after all, they are different functions. However, $f = gh$ if $h : X \to X$ is the bijection given by $h(a) = a$, $h(b) = c$, and $h(c) = b$; therefore f and g satisfy definition (2). The bijection h rearranges the set $\{a, b, c\}$, eliminating the discrepancy between the two functions due to the labeling of the elements of the domain. If i is the bijection given by $i(x) = x$, $i(y) = z$, and $i(z) = y$, then $f = ig$; therefore f and g satisfy definition (3). It is now clear that the functions f and g of Figure 2.4 are equivalent according to definitions (2), (3), and (4).

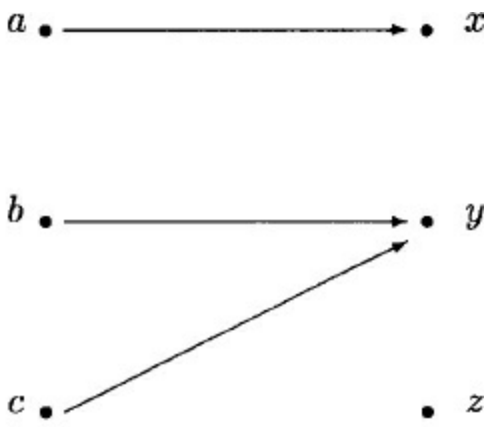
EXAMPLE 2.10

Are the functions f and g of Figure 2.5 equivalent?

Figure 2.5 Equivalent functions when domain and codomain are unlabeled.



f



g

Solution: The functions f and g are equivalent only when both X and Y are unlabeled, i.e., according to definition (4). Define $h: X \to X$ by $h(a) = c$, $h(b) = b$, and $h(c) = a$, and define $i: Y \to Y$ by $i(x) = z$, $i(y) = x$, and $i(z) = y$. Then $f = igh$.

The domain and codomain of a function may be labeled or unlabeled sets, leading to four types of functions to be counted. Furthermore, functions may be classified according to whether they are one-to-one, onto, both, or not necessarily either. Altogether there are 16 cases, as shown in Table 2.2.

Table 2.2 The number of functions from X to Y, where $|X| = m$ and $|Y| = n$.

X \ Y	labeled		unlabeled		
labeled	total	n^m	total	$B(m)$	$m \le n$
	1–1	$P(n,m)$	1–1	1	$m \le n$
				0	$m > n$
	onto	$T(m,n)$	onto	$\left\{ {m \atop n} \right\}$	
	bijections	$m!\delta(m,n)$	bijections	$\delta(m,n)$	
unlabeled	total	$\binom{m+n-1}{m}$	total	$p(m)$	$m \le n$
	1–1	$\binom{n}{m}$	1–1	1	$m \le n$
				0	$m > n$
	onto	$\binom{m-1}{n-1}$	onto	$p(m,n)$	
	bijections	$\delta(m,n)$	bijections	$\delta(m,n)$	

We assume that $f: X \to Y$ is a function from a set X with m elements to a set Y with n elements. Let

us consider the entries in the table, beginning with the X labeled, Y labeled box.

X labeled, Y labeled

We have already noted that the total number of such functions is n^m.

As for the second entry, every one-to-one (1–1) function corresponds to an ordered selection of m objects from the set Y. There are n choices for the first object selected, $n - 1$ choices for the second, and so on, leading to the formula

$$P(n, m) = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}, \quad m \le n.$$

If $m > n$, there are no one-to-one functions so we set $P(n, m) = 0$.

In the case of bijections, we either have two sets of the same cardinality or we don't. If $m \neq n$, then no bijection is possible. However, if $m = n$, then there are $m!$ ways to match up the two sets. We define $\delta(m, n)$ to be 0 if $m \neq n$ and 1 if $m = n$. Thus, the number of bijections is $m!\delta(m, n)$. If either X or Y (or both) is unlabeled, then all bijections look the same, so the total number is $\delta(m, n)$.

X unlabeled, Y labeled

A one-to-one function $f : X \to Y$ is equivalent to a selection of m elements of Y without regard to order. The number of such selections is the value of the binomial coefficient

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

A function $f : X \to Y$ is equivalent to a distribution of m identical, elements (the elements of X) into n different boxes (the elements of Y). A box may receive any number of objects, including zero. Suppose we represent the m identical objects with m copies of the symbol *. The n boxes are represented by a set of $n - 1$ vertical lines|. The placement of the objects in the boxes is indicated by a linear ordering of the *'s and |'s. For example, * * * | * || * means that the first box (to the left of the first |) contains three objects, the second box contains one object, the third contains no objects, and the fourth contains one object. The total number of items in the linear ordering is $m + n - 1$, and n of these are |'s. Therefore, the total number of functions is

$$\binom{m+n-1}{n-1}.$$

If f is onto, then each box must contain at least one object, and there are $m - n$ objects to distribute freely. Hence, the number of onto functions is

$$\binom{m-n+n-1}{n-1} = \binom{m-1}{n-1}.$$

Can you furnish a more direct combinatorial proof of this formula?

X labeled, Y unlabeled

If X is labeled and Y unlabeled, then the total number of functions from X to Y is the number of ways X may be partitioned into unlabeled parts (corresponding to the images under f). We prefer to define a formula only when $m \le n$, in which case the magnitude of n becomes unimportant (X can't be divided

into more than m parts). The Bell number $B(m)$ is the number of such functions.

If $m > n$, then one-to-one functions do not exist. However, if $m \leq n$, then all one-to-one functions look alike when X is unlabeled. This observation accounts for the 1–1 entries of the X unlabeled, Y labeled and X unlabeled, Y unlabeled boxes.

The number of onto functions when X is labeled and Y is unlabeled is $\left\{ {m \atop n} \right\}$, a Stirling number of the second kind. Clearly,

$$\left\{ {m \atop n} \right\} = \frac{T(m,n)}{n!},$$

as dividing by $n!$ permutes the labels of the n sets into which X has been divided.

X unlabeled, Y unlabeled

The total number of functions when X and Y are unlabeled and $m \leq n$ is denoted $p(m)$, and the values of $p(m)$ are the *partition numbers*. The number of onto functions is the partition number $p(m, n)$.

Four relations are obtained by comparing the total number of functions in each box of Table 2.2 to the number of onto functions. Thus

$$n^m = \sum_{j=1}^{m} T(m,j) \binom{n}{j} \tag{2.27}$$

$$\binom{m+n-1}{n-1} = \sum_{j=1}^{m} \binom{m-1}{j-1} \binom{n}{j} \tag{2.28}$$

$$B(m) = \sum_{j=1}^{m} \left\{ {m \atop j} \right\} \tag{2.29}$$

$$p(m) = \sum_{j=1}^{m} p(m,j). \tag{2.30}$$

The relation (2.27) follows from the fact that every function $f: X \to Y$ is onto *some* subset $Y' \subseteq Y$ of cardinality j, where $1 < j \leq m$. The binomial coefficient $\binom{n}{j}$ "chooses" Y' and $T(m, j)$ counts the number of functions from X onto Y'. Relations (2.28) through (2.30) are proved similarly.

EXERCISE

2.41 Verify the formulas for the X unlabeled, Y labeled case in Table 2.2.

2.42 Let $|X| = m$, $|Y| = n$.

(a) How many possible relations are there from X to Y?

(b) How many relations are there if X is unlabeled and Y is labeled?

(c) How many relations are there if X is labeled and Y is unlabeled?

(d) How many relations are there if both X and Y are unlabeled?

Hint: In each case, the relation $R: X \to Y$ may be viewed as a function $f: X \to \mathrm{P}(Y)$, defined by $f(x) = R(x)$. Now use the techniques associated with the fundamental counting problem for functions. For example, the answer for (b) is $\binom{m+2^n-1}{m}$.

2.43 Let $|X| = m$. An *algebra* on X is a subset S of $P(X)$ with the following properties:

1. $X \in S$.

2. $A \in S$ implies $X - A \in S$.

3. $A \in S$ and $B \in S$ imply $A \cup B \in S$.

(a) How many algebras on X are possible if X is labeled?

(b) How many if X is unlabeled?

2.44 Let f be a random function from $\{1,\ldots,n\}$ to $\{1,\ldots,n\}$ and let $r(n)$ be the expected number of elements in the range of f. Find $\lim_{n\to\infty} r(n)/n$.

2.45 How many onto functions $f : X \to X$, $|X| = m$ (X labeled), have the property that $f(f(x)) = x$ for all $x \in X$? Find a recurrence relation or an explicit formula. See Exercises 2.30 and 2.69.

2.46 Give a combinatorial proof of the formula for the number of onto functions from an unlabeled set to a labeled set.

2.5 Counting with symmetry

We have classified functions $f : X \to Y$ ($|X| = m$, $|Y| = n$), where X and Y are labeled or unlabeled sets, and we enumerated these functions. Now we generalize the notion of labeled and unlabeled sets. For instance, recall that two functions $f: X \to Y$ and $g : X \to Y$ are equivalent in the X unlabeled, Y labeled sense if there exists a bijection $h: X \to X$ such that $f = gh$. The bijection h can be viewed as a permutation of X (in fact, *any* permutation in the symmetric group S_m.). What happens if we restrict the permutations to a specified subgroup G of S_m? If G is the identity group (e), for example, then we obtain the X labeled case; while if $G = S_m$, we obtain the X unlabeled case. Nontrivial subgroups G give rise to interesting intermediate cases. In these cases, Pólya's theorem for the number of inequivalent functions allows us to count quite complicated configurations, including nonisomorphic graphs. (See Section 3.3 for definitions about graphs.) More generally, if one group G acts on X and another group H acts on Y, then the number of inequivalent functions is given by de Bruijn's formula, which enumerates more intricate structures such as self-complementary graphs.

A *group* G is a nonempty set on which is defined a binary operation * satisfying the following three laws:

1. For all $x, y, z \in G$, $x * (y * z) = (x * y) * z$ (associativity).

2. There exists an element $e \in G$ with the property that, for all $x \in G$, $x * e = e * x = x$.

3. For every $x \in G$, there exists an element $x^{-1} \in G$ with the property that $x * x^{-1} = x^{-1} * x = e$.

The element e is called the *identity* of G. The element x^{-1} is called the *inverse* of x. One can easily prove that the identity element of a group is unique and that the inverse x^{-1} of each element x is unique.

We sometimes indicate a group by an ordered pair, e.g., $(G, *)$, to emphasize both the set and the operation. We usually suppress the group operation sign and write xy for $x * y$. We abbreviate xx by x^2, $x^{-1}x^{-1}$ by x^{-2}, etc. For any $x \in G$, we set $x^0 = e$.

EXAMPLE 2.11

The set $\mathbf{Z}$ is a group with respect to addition.

A *finite group* is a group with a finite number of elements, and the *order* of a finite group is the number of elements in it.

The *cyclic group* $\mathbf{Z}_n$, of order n, is the set $\{0,\ldots,n-1\}$ under the clock addition operation * defined by $x * y = x + y$ if $x + y < n$ and $x*y = x + y - n$ if $x + y \geq n$.

A group G is *abelian* if $xy = yx$ for all $x, y \in G$. Otherwise, G is *nonabelian*.

EXAMPLE 2.12

The group $\mathbf{Z}_n$, is abelian.

The *order* of an element $x \in G$ is the least positive integer n for which $x^n = e$. If there is no such integer, then we say that x has *infinite order*.

EXAMPLE 2.13

In $\mathbf{Z}_4$, the elements 0, 1, 2, 3 have orders 1, 4, 2, 4, respectively.

The *symmetric group* S_n consists of the $n!$ permutations of an n-set (e.g., N_n). The group operation * is composition of permutations. The elements of S_n are conveniently written in *cycle notation*. Thus,

$$(1\ 2\ 3)(4\ 8)(3\ 6\ 7)(5)(9)$$

is the element of S_9 which maps 1 to 2 to 3 to 1, *transposes* 4 and 8, maps 3 to 6 to 7 to 3, and *fixes* 5 and 9. To multiply two permutations together, we just compute the result of the composition of the two bijections (reading left to right). For example,

$$(1\ 2\ 3)(4\ 5) * (1\ 2\ 3\ 4\ 5) = (1\ 3\ 2\ 4)(5).$$

Since $(1\ 2)(1\ 3) \neq (1\ 3)(1\ 2)$, the symmetric group S_n, is nonabelian for all $n \geq 3$.

A *homomorphism* from one group $(G_1, *_1)$ to another group $(G_2, *_2)$ is a map $f : G_1\ G_2$ which preserves multiplicative structure: $f(x *_1 y) = f(x) *_2 f(y)$ for all $x, y \in G_1$. If the homomorphism is a bijection we call it an *isomorphism* and say that the two groups G_1 and G_2 are *isomorphic*; we write $G_1 \simeq G_2$. A one-to-one homomorphism is called a *monomorphism* and an onto homomorphism is called an *epimorphism*. An isomorphism from a group G to itself is called an *automorphism* of G.

Suppose that G_1 and G_2 are two groups. The *product* of G_1 and G_2, denoted $G_1 \times G_2$, is the set of ordered pairs $\{(g_1, g_2) : g_1 \in G_1, g_2 \in G_2\}$ subject to the multiplication rule $(g_1, g_2) * (g'_1, g'_2) = (g_1 g'_1, g_2 g'_2)$.

EXAMPLE 2.14

The product $\mathbf{Z}_2 \times \mathbf{Z}_2$ is a four-element group. It is not isomorphic to $\mathbf{Z}_4$, for $\mathbf{Z}_2 \times \mathbf{Z}_2$ has three elements of order 2 and $\mathbf{Z}_4$ has only one.

We say that H is a *subgroup* of G if H is a subset of G and H is a group with respect to the group operation of G. We say that H is a *normal subgroup* of G if it is invariant under conjugation by

elements of G, that is, $xHx^{-1} = H$ for all $x \in G$.

EXAMPLE 2.15

The two-element group $\{(1\ 2)(3), (1)(2)(3)\}$ is a subgroup of the six-element group S_3. It is not a normal subgroup. Why not?

The symmetric group S_n is especially important because every finite group is isomorphic to a subgroup of some S_n. This is Cayley's theorem.

Cayley's theorem (1854). If G is a finite group of order n, then G is isomorphic to a subgroup of S_n.

Proof. For each element $g \in G$, we define a function f_g: $G \to G$ by the rule $f_g(a) = ag$ (right multiplication by g). Note that f_g is a bijection because it has an inverse, namely, $f_{g^{-1}}$. We check: $(f_g \circ f_{g^{-1}})(a) = ag^{-1}g = a$ and $(f_{g^{-1}} \circ f_g)(a) = agg^{-1} = a$. Since each f_g is a permutation of the n-set G, we can define a map ϕ: $G \to S_n$ by $\phi(g) = f_g$. We claim that ϕ is a monomorphism. First we check that ϕ is an isomorphism: $\phi(gh)(a) = f_{gh}(a) = a(gh) = (ag)h = f_h(f_g(a)) = (f_g f_h)(a) = (\phi(g)\phi(h))(a)$. Now we check that ϕ is one-to-one: If $\phi(g)(a) = \phi(h)(a)$, then $f_g(a) = f_h(a)$, which implies that $ag = ah$, and $g = h$.

The *dihedral group* D_n, of order $2n$, consists of the set of symmetries of a regular n-gon. If we number the vertices of the n-gon $1,\ldots,n$, then we see that D_n is a subgroup of S_n. Specifically, the subgroup is generated by two permutations: the *rotation* $r = (1\ 2\ 3\ 4\ \ldots\ n)$ and a *flip* f along an axis of symmetry of the n-gon. If n is odd we take the flip to be $f = (n)(1\ n-1)(2\ n-2)(3\ n-3)\ \ldots\ (\frac{n-1}{2}\ \frac{n+1}{2})$. If n is even we choose $f = (1\ n-1)(2\ n-2)\ldots(\frac{n}{2}-1\ \frac{n}{2}+1)$. Each element of D_n can be written in the form $r^a f^b$, where $a \in \{1, n-1\}$ and $b \in \{0, 1\}$. Two such elements are multiplied using the basic rules $r^n = e$, $f^2 = e$, and $rf = fr^{-1}$. From these basic rules it is possible to compute all other products. We say that D_n has the *presentation* $\langle r, f : r^n = 1, f^2 = 1, rf = fr^{-1}\rangle$. For information about group presentations the reader should consult [17].

We have noted that every element of S_n can be expressed as a product of cycles. A cycle of length 2 is called a *transposition* and a cycle of length 1 is called *a fixed point*. Cycles of length greater than 2 can be written as products of transpositions. For example, $(1\ 2\ 3) = (1\ 2)(1\ 3)$. A permutation may be written as a product of transpositions and fixed points in more than one way. However, the number of transpositions is always even or always odd. A permutation is accordingly called an *even* permutation or an *odd* permutation. It follows from the first homomorphism theorem for groups, to be discussed shortly, that S_n contains $n!/2$ even permutations and $n!/2$ odd permutations. (This fact follows more simply from the observation that $f(\sigma) = (12)\sigma$ is a bijection between the set of even permutations of S_n and the set of odd permutations of S_n.) As the identity permutation is even and the even permutations are closed under multiplication and taking inverses, the even permutations are a group.

The *alternating group* A_n is the group of even permutations of the set $\{1,\ldots,n\}$. The group has

order $\frac{1}{2}n!$.

Let G be a group (with identity element e) and X a nonempty set. An *action* of G on X is a function θ: $G \times X \to X$ which satisfies the following two conditions:

1. For every $x \in X$, we have $\theta(e, x) = x$.

2. For every $g, h \in G$ and $x \in X$, we have $\theta(g, \theta(h, x)) = \theta(gh, x)$.

For convenience, we write $\theta(g, x)$ as gx, so that the two conditions become:

1. $ex = x$.
2. $g(hx) = (gh)x$.

Remember, however, that g and h are group elements and x is a set element.

We note that each element $g \in G$ yields a permutation of the set X (defined by sending x to gx). For if $gx = gy$, then $x = y$ (hence the map is one-to-one) and $gg^{-1}x = x$ (hence the map is onto).

EXAMPLE 2.16

The symmetric group S_n, acts on N_n, by the *natural* action $gx = g(x)$, where $g(x)$ is the image of x under the function g: $N_n \to N_n$.

EXAMPLE 2.17

The cyclic group Z_n, acts on N_n by

$$gx = \begin{cases} g + x & \text{if } g + x \leq n \\ g + x - n & \text{if } g + x > n. \end{cases}$$

Here g denotes the equivalence class representative of $[g]$ that lies between 1 and n.

If $x \in X$, then the *orbit* of x (under the action θ) is $\text{orb}(x) = \{gx: g \in G\}$. Orbits constitute equivalence classes of X; that is, X is partitioned into orbits. If there is only one orbit, then the action θ is *transitive*. Both examples above are transitive actions.

To find the size of orb(x), note that $gx = hx$ if and only if $h^{-1}gx = x$. Let $G_x = \{g \in G : gx = x\}$. We call G_x the *stabilizer* of x. We leave it to the reader to check that G_x is a normal subgroup G. Now $gx = gy$ if and only if $g \in hG_x$, that is, if g and h are elements of the same coset of G_x. Therefore, the number of distinct values of gx is the number of cosets of Gx:

$$|\text{orb}(x)| = [G : G_x] = \frac{|G|}{|G_x|}. \tag{2.31}$$

Let G be a finite group. For each $g \in G$, define f_g: $G \to G$ by $f_g(x) = gxg^{-1}$. We say that G acts on itself by *conjugation*. The stabilizer of an element $x \in G$ is called the *centralizer* of x and is denoted $C(x)$. We call orb(x) the *conjugacy class* of x, and denote it by ccl(x). For the conjugacy action, we have

$$|\text{ccl}(x)| = [G : C(x)] = \frac{|G|}{|C(x)|}. \tag{2.32}$$

Two permutations are in the same conjugacy class if and only if they have the same cycle structure.

Therefore, the number of conjugacy classes equals the partition number $p(n)$, where $n = |G|$.

Burnside's lemma. Let G be a finite group acting on a finite nonempty set X. For each $g \in G$, let f_g be the number of elements of X fixed by g. Then number of orbits n is given by

$$n = \frac{1}{|G|} \sum_{g \in G} f_g.$$

Proof. The proof is a nice example of the technique of enumerating a set two different ways and comparing the results. The set is

$$S = \{(g, x) : g \in G,\ x \in X, \text{ and } gx = x\}.$$

On the one hand, by definition of f_g, we have $|S| = \Sigma_{g \in G} f_g$. On the other hand, counting from the perspective of the elements of X and letting x' denote orbit representatives, we have

$$\begin{aligned} |S| &= \sum_{x \in X} |G_x| \\ &= \sum_{x \in X} \frac{|G|}{|\mathrm{orb}(x)|} \\ &= |G| \sum_{x \in X} \frac{1}{|\mathrm{orb}(x)|} \\ &= |G| \sum_{x'} \frac{|\mathrm{orb}(x')|}{|\mathrm{orb}(x')|} \\ &= |G| \sum_{x'} 1 \\ &= |G| n. \end{aligned}$$

Equating the two expressions for $|S|$, we obtain $\sum_{g \in G} f_g = |G| n$, from which the desired relation follows instantly.

EXAMPLE 2.18 Average number of fixed points of a permutation

How many fixed points has the average permutation of $\{1, 2, 3, \ldots, n\}$? Recall that we solved this problem using probability in Example 1.27.

Solution: Applying Burnside's lemma to the natural action of S_n, on N_n (which is transitive), we obtain

$$\frac{1}{n!} \sum_{g \in S_n} f_g = 1.$$

This is the average number of fixed points.

We now assume that $f : X \to Y$ is a function from a set X of m elements to a set Y of n elements. In the terminology of Pólya's theory of counting, the elements of Y are *labels*, and f is a *labeling* of X. Suppose that a finite group G acts on X. We picture this situation with the following diagram:

$$f:\ \overset{G}{\downarrow}\ X \longrightarrow Y$$

This action of G on X induces an action of G on the set of labelings Y^X as follows: $(gf)(x) = f(g^{-1}x)$ for all $x \in X$. We check the two axioms for an action:

1. $(ef)(x) = f(e^{-1}x) = f(x)$ for all $x \in X$.

2. $(g(hf)(x) = (hf)(g^{-1}x) = f(h^{-1}x) = f((gh)^{-1}x) = ((gh)f)(x)$ for $x \in X$.

The set Y^X of functions is partitioned into equivalence classes by this action. Functions in different equivalence classes are called *G-inequivalent functions*. By definition, the number of G-inequivalent functions is the number of orbits of the action of G on Y^X.

Number of orbits theorem. When G (finite) acts on the set of functions Y^X ($|X| = m$, $|Y| = n$), the number of orbits is

$$\frac{1}{|G|}\sum_{g\in G} n^{c(1)+c(2)+\cdots+c(m)}, \tag{2.33}$$

where $c(i)$ is the number of cycles of length i of g (regarded as a permutation of X).

Proof. Suppose that g, when regarded as a permutation of X, has *cycle structure* $c = (c(1), c(2), \ldots, c(m))$. The functions fixed by g are precisely those which are constant on each cycle. As there are $c(1) + \cdots + c(m)$ cycles, each of which may be assigned one of n images, the number of functions fixed by g is $f_g = n^{c(1)+\cdots+c(m)}$. Our conclusion now follows directly from Burnside's lemma.

EXAMPLE 2.19 Stacks of chips

How many stacks of 11 poker chips are possible with two colors of poker chips (red and white)?

Solution: Let X be the set of 11 positions in the stack and $Y = \{\text{red, white}\}$. The group $S_2 = \{e, \tau\}$ acts on X, the identity e leaving the stack alone and τ turning the stack upside down. We need to know the cycle structures of e and τ. Certainly, e consists of 11 fixed points, so $c(1) = 11$ and $c(i) = 0$ for $2 \le i \le 11$. And τ consists of one fixed point (the middle poker chip) and five transpositions, so $c(1) = 1$, $c(2) = 5$, and $c(i) = 0$ for $3 \le i \le 11$. Therefore, the number of S_2-inequivalent stacks is

$$\frac{1}{2}(2^{11} + 2^6) = 1056.$$

EXAMPLE 2.20 Circular necklaces

How many 10-bead circular necklaces may be made using black or white beads?

Solution: Symmetries of the circular arrangement of beads are of six types: identity, rotations with one cycle, rotations with two cycles, rotations with five cycles, reflections through two opposite beads, and reflections through two opposite "sides?' The number of necklaces is

$$\frac{1}{20}(1\cdot 2^{10} + 4\cdot 2^1 + 4\cdot 2^2 + 2^5 + 5\cdot 2^6 + 5\cdot 2^5) = 78.$$

While it is possible at this point to enumerate some rather complicated structures (nonisomorphic graphs, for instance), we prefer to do so only after developing some additional machinery—cycle

indexes—in the next section.

EXERCISE

2.47 (a) How many conjugacy classes has S_5?

(b) Let $x = (1\ 2\ 3)(4)(5)$. Find $|\mathrm{ccl}|(x)|$ and $C(x)$.

2.48 Let $\mathbf{Z}_4$ rotate a cube around an axis passing through the centers of opposite faces. Verify Burnside's lemma for

X_1 = the set of vertices of the cube.

X_2 = the set of faces of the cube, and

X_3 = the set of edges of the cube.

2.49 (*a*) What is the average number of 1-cycles in the group A_n?

(b) What about in D_n?

2.6 Cycle indexes

As in the previous section, we let $c = (c(1),\ldots,c(m))$ be the cycle structure of $g \in G$ when G acts on X (G finite, $|X| = m$). We assign to g the monomial

$$x_1^{c(1)} x_2^{c(2)} \cdots x_m^{c(m)},$$

where the x_i are variables in a commutative ring containing the rational numbers. The *cycle index* $Z(G)$ is the average of these monomials:

$$Z(G) = \frac{1}{|G|} \sum_{g \in G} x_1^{c(1)} x_2^{c(2)} \cdots x_m^{c(m)}. \tag{2.34}$$

The cycle index $Z(G)$ contains complete information about the cycle lengths of the various permutations g of the group action. George Pólya (1887–1985) chose the letter Z to stand for the German word *Zycklenzeiger* ("cycle indicator"). As Pólya said, "The cycle index knows many things." For instance, letting each $x_i = n$ (a substitution denoted $Z(G)[x_i \leftarrow n]$), we obtain the formula for the number of G-inequivalent functions in Y^X ($|X| = m$, $|Y| = n$):

$$Z(G)[x_i \leftarrow n] = \frac{1}{|G|} \sum_{g \in G} n^{c(1)+\cdots+c(m)}. \tag{2.35}$$

We next determine the cycle indexes of the most important group actions: E_n, $\mathbf{Z}_n$, D_n, S_n, A_n. The set acted upon is always $X = \{1,\ldots n\}$.

The identity group E_n. The identity group consists of only the identity element e, which fixes every element of X. There are n 1-cycles, and the cycle index is

$$Z(E_n) = x_1^n. \tag{2.36}$$

The cyclic group $\mathbf{Z}_n$. Given $g \in G$, $x \in X$, the length of the cycle containing x is the minimum positive k for which $gk + x \equiv x \pmod{n}$, or $gk \equiv 0 \pmod{n}$. Because k is independent of x, all cycles of g have length k. If g contains j cycles, then $jk = n$, from which it follows that k is a divisor of n and $j = n/k$.

To determine the number of elements g corresponding to each value of k, observe that $k'n/k$ has order k whenever $\gcd(k', k) = 1$. There are $\phi(k)$ such values of k', by definition of Euler's ϕ-function. As $\Sigma_{k|n}\, \phi(k) = n$, there are exactly $\phi(k)$ values of g for each k/n. Therefore

$$Z(\mathbf{Z}_n) = \frac{1}{n} \sum_{k|n} \phi(k) x_k^{n/k}. \tag{2.37}$$

The dihedral group D_n. As D_n contains Z_n as a subgroup, the cycle index of D_n will contain all the terms in the formula (2.37). The other elements of D_n are "flips" (reflections). If n is odd, each flip fixes one element of X and contains $(n-1)/2$ transpositions. If n is even, half of the flips contain $n/2$ transpositions and half contain two fixed points and $(n-2)/2$ transpositions. Putting these facts together, we obtain the formulas

$$Z(D_n) = \frac{1}{2} Z(\mathbf{Z}_n) + \begin{cases} \frac{1}{2} x_1 x_2^{(n-1)/2} & (n \text{ odd}) \\ \frac{1}{4} \left(x_2^{n/2} + x_1^2 x_2^{(n-2)/2} \right) & (n \text{ even}). \end{cases} \tag{2.38}$$

The symmetric group S_n. A permutation $g \in S_n$ can have any cycle structure $c = (c(1),\ldots,c(n))$, where

$$1c(1) + 2c(2) + \cdots + nc(n) = n. \tag{2.39}$$

The number of solutions to this equation is a good counting puzzle in its own right. Solutions may be generated by ordering the elements of X and partitioning the elements from left to right into cycles of the appropriate lengths. Each of the $n!$ orderings of X gives rise to repeated solutions due to interchanging cycles and writing cycles down in more than one way. Because there are $c(k)!$ ways to list cycles of length k and each cycle may be written k ways, the number of solutions is

$$h(c) = \frac{n!}{\prod_{k=1}^{n} k^{c(k)} c(k)!}. \tag{2.40}$$

The cycle index of the symmetric group is

$$Z(S_n) = \frac{1}{n!} \sum_{c} h(c) x_1^{c(1)} \cdots x_n^{c(n)}. \tag{2.41}$$

The alternating group A_n. The alternating group consists of the $\frac{1}{2}n!$ even permutations of S_n. When a permutation is written as a disjoint product of cycles, it is easy to tell whether it is even or odd. Because each odd cycle is equal to the product of an even number of transpositions, the number of odd cycles has no bearing on whether a permutation is even. However, each even cycle is equal to the product of an odd number of cycles. Hence, for a permutation to be even, it must be composed of an even number of disjoint cycles of even length. Therefore,

$$\frac{1}{2} \left(1 + (-1)^{c(2)+c(4)+\cdots+c(2\lfloor n/2 \rfloor)} \right)$$

counts 1 for every even permutation and 0 for every odd permutation in S_n. This establishes the cycle index for the alternating group:

$$Z(A_n) = \frac{1}{n!} \sum_{c} h(c) \left(1 + (-1)^{c(2)+c(4)+\cdots+c(2\lfloor n/2 \rfloor)} \right) x_1^{c(1)} \cdots x_n^{c(n)}. \tag{2.42}$$

EXAMPLE 2.21 Circular necklaces (again)

How many 11-bead circular necklaces can be made with two types of beads?

Solution: The appropriate group is D_{11} acting on the set $X = \{1,\ldots,11\}$. Hence

$$\begin{aligned} Z(D_{11}) &= \frac{1}{2}Z(\mathbf{Z}_{11}) + \frac{1}{2}x_1x_2^5 \\ &= \frac{1}{22}\sum_{k|11}\phi(k)x_k^{n/k} + \frac{1}{2}x_1x_2^5. \end{aligned}$$

The only divisors of 11 are 1 and 11, for which $\phi(1) = 1$ and $\phi(11) = 10$. Thus

$$Z(D_{11}) = \frac{1}{22}(x_1^{11} + 10x_{11}) + \frac{1}{2}x_1x_2^5.$$

The number of necklaces that can be made with two types of beads is obtained by making the substitution

$$\begin{aligned} Z(D_{11})[x_i \leftarrow 2] &= \frac{1}{22}(2^{11} + 10 \cdot 2) + \frac{1}{2} \cdot 2 \cdot 2^5 \\ &= 126. \end{aligned}$$

Recall that in Example 2.19 we determined that there are 1056 different stacks of 11 poker chips using two types of chips. The number of necklaces with 11 beads of two types is smaller because a necklace has more symmetries than a stack of chips.

EXERCISE

2.50 Calculate $Z(\mathbf{Z}_3)$, $Z(D_3)$, $Z(A_3)$, and $Z(S_3)$.

2.51 How many 10-bead circular necklaces may be made using three types of beads?

2.52 Calculate the number of 100-bead circular necklaces that may be made using two types of beads.

2.53 How many 100-bead circular necklaces can be made with 50 white beads and 50 black beads?

Hint: If you don't see how to solve this now, consult the next section.

2.54 How many ways may the eight vertices of a cube be colored with three colors (up to symmetry of the cube)? How many ways may the six faces be colored? How many ways may the 12 edges be colored?

2.7 Pólya's theorem

Pólya's theorem. Let G act on X and therefore on the set of functions $f: X \to Y$ (G finite, $|X| = m$, $|Y| = n$). Suppose that $F(y(1),\ldots,y(n))$ is the set of G-inequivalent functions for which $|f^{-1}(y_i)| = y(i)$, where $1 \le i \le n$. Then

$$Z(G)\left[x_i \leftarrow \sum_{j=1}^{n} y_j^i\right] = \sum |F(y(1),\ldots,y(n))| y_1^{y(1)} \cdots y_n^{y(n)}, \tag{2.43}$$

where the right-hand sum is taken over all n-tuples $(y(1),\ldots,y(n))$ with

$$\sum_{j=1}^{n} y(j) = m.$$

Proof. We need to show that the coefficient of ${y^{y(1)}}_1 \cdots {y^{y(n)}}_n$ on the left side of the relation (2.43) is $|F(y(1),\ldots,y(n))|$. By Burnside's lemma,

$$|F(y(1),\ldots,y(n))| = \frac{1}{|G|}\sum_{g\in G} f_g, \tag{2.44}$$

where f_g is the number of functions in $F(y(1),\ldots,y(n))$ fixed by g.

Suppose that g has cycle structure $c = (c(1),\ldots,c(n))$. If the function

$$f \in F(y(1),\ldots,y(n))$$

is fixed by g, then f is constant on each cycle of g; that is, each cycle of g lies entirely within one of the inverse images $f^{-1}(y_i)$. Picture an $m \times 1$ box (the elements of X) partitioned into n sections of sizes $y(1),\ldots,y(n)$ (the inverse images of f). Then f_g is the number of possible packings of this box with $c(i)$ blocks of sizes i, where $1 \le i \le m$ (the cycles of g). The polynomial on the left side of the relation (2.43) is

$$\frac{1}{|G|}\sum_{g\in G} x_1^{c(1)}\cdots x_m^{c(m)}\left[x_i \leftarrow \sum_{j=1}^{n} y_j^i\right] = \frac{1}{|G|}\sum_{g\in G}\prod_{k=1}^{m}\left(\sum_{j=1}^{n} y_j^k\right)^{c(k)}.$$

Let us consider how the $y_1^{y(1)}\cdots y_n^{y(n)}$ terms are formed in the relation above. Suppose that each term $(\sum_{j=1}^{n} y_j^i)^{c(i)}$ is expanded as a product of $c(i)$ factors. Then each term in the product of these multiplicands is obtained by choosing a y^i_i from each factor. The contribution to $y_1^{y(1)}\cdots y_n^{y(n)}$ is the number of ways the exponents ($c(i)$ units of size i) may be arranged to equal $y(i)$ for each $1 \le i \le n$. These arrangements are clearly equivalent to the box packings described above. Therefore, the coefficient of $y_1^{y(1)}\cdots y_n^{y(n)}$ is

$$\frac{1}{|G|}\sum_{g\in G} f_g = |F(y(1),\ldots,y(n))|,$$

as we needed to show.

EXAMPLE 2.22 Circular necklaces (yet again)

Recall Example 2.20, which counted the number of circular necklaces with 10 beads, using black or white beads. How many such necklaces are made of seven white beads and three black beads?

Solution: The cycle index is

$$Z(\mathbf{D}_{10}) = \frac{1}{20}(x_1^{10} + 4x_{10} + 4x_5^2 + x_2^5 + 5x_1^2x_2^4 + 5x_2^5).$$

Making the substitution $x_i \leftarrow +y^i_1 + y^i_2$ yields

$$y_1^{10}+y_1^9y_2+5y_1^8y_2^2+8y_1^7y_2^3+16y_1^6y_2^4+16y_1^5y_2^5+16y_1^4y_2^6+8y_1^3y_2^7+5y_1^2y_2^8+y_1y_2^9+y_2^{10}.$$

We see that there are eight necklaces with seven white beads and three black beads. Can you sketch these eight necklaces?

EXAMPLE 2.23 Circular necklaces (one more time)

Making the substitution $x_i \leftarrow a^i + 1$ into $Z(D_{11})$, we obtain the polynomial

$$a^{11} + a^{10} + 5a^9 + 10a^8 + 20a^7 + 26a^6 + 26a^5 + 20a^4 + 10a^3 + 5a^2 + a + 1.$$

(We use the term $a^i + 1$ instead of $y^i_1 + y^i_2$ because it is simpler.) This polynomial tells us the number of 11-bead circular necklaces made of two types of beads, according to the number of beads of each type. For instance, there are 20 necklaces with seven white beads and four black beads.

EXERCISE

2.55 How many stacks of 10 poker chips contain four black chips and six white chips (up to inversion of the stack)?

2.56 How many circular necklaces of 10 beads may be made from two types of beads with five of each type?

2.57 How many 11-bead circular necklaces are there with three red beads, three white beads, and five blue beads?

2.58 How many ways may eight identical markers be placed on an 8 × 8 square grid (up to a rotation of the grid)?

2.59 How many ways may the vertices of an octagon be colored with three colors (up to symmetry of the octagon)?

2.60 How many ways may the faces of a cube be colored using three colors (up to symmetry of the cube)?

2.61 (a) How many ways may the 12 faces of a regular dodecahedron be colored with two colors?

(b) How many ways with six faces of each color?

2.62 How many ways can the faces of a regular icosahedron be colored with three colors (up to symmetry of the icosahedron)?

2.8 The number of graphs

In order to use Pólya's theorem to count nonisomorphic graphs of order n, we need to determine the cycle index of the appropriate group action. A graph $G = (V, E)$ may be identified with a

function $f : [V]^2 \to \{0, 1\}$, where $f(\{x, s\}) = 1$ or 0 according to whether or not $\{x, y\}$ is a member of E. We say that two graphs G_1 and G_2 are *isomorphic* if there is a bijection between their vertex sets V_1 to V_2 that preserves adjacency. We normally regard two isomorphic graphs as the same graph. Two graphs are isomorphic if the corresponding functions are equal up to a permutation of V. As any permutation of V is allowed, the group acting on V is S_n, where $n = |V|$. This group induces an action on $[V]^2$ by the rule $\{x, y\}$ $\{gx, gy\}$. We call this action $[S_n]^2$. Our goal is to calculate the cycle index $Z([S_n]^2)$.

Assume that $g \in S_n$ has cycle structure $c = (c(1),\ldots,c(n))$. We determine the cycle structure of g as a permutation of $[V]^2$ by assuming that $\{x, y\} \in [V]^2$ and examining four cases. We call the cycles in $[V]^2$ *pair-cycles*.

Case 1. Suppose that x and y are in cycles of different lengths a and b. There are $c(a)$ choices for which cycle contains x and $c(b)$ choices for which cycle contains y. Given these choices, the ab ordered pairs in the Cartesian product of the two cycles are partitioned into pair-cycles of length lcm(a, b). The number of such cycles is ab/lcm(a, b) = gcd(a, b). The contribution to $Z([S_n]^2)$ is

$$\prod_{1 \le a < b \le n} x_{\mathrm{lcm}(a,b)}^{c(a)c(b)\gcd(a,b)}. \tag{2.45}$$

Case 2. Suppose that x and y are in different cycles of the same length a. There are $\binom{c(a)}{2}$ choices for the cycles. The pair-cycle has length a, and the number of pair-cycles is a. The contribution to $Z([S_n]^2)$ is

$$\prod_{1 \le a \le n} x_a^{a\binom{c(a)}{2}}. \tag{2.46}$$

Case 3. Suppose that x and y are in the same cycle of odd length a. There are $c(a)$ choices for the cycle containing x and y and $(a - 1)/2$ choices for the gap between x and y. The pair-cycles all have length a. The contribution to $Z([S_n]^2)$ is

$$\prod_{a \text{ odd}} x_a^{c(a)(a-1)/2}. \tag{2.47}$$

Case 4. Suppose that x and y are in the same cycle of even length a. This case is the same as Case 3 except for one important difference. There are still $c(a)$ choices for the cycle containing x and y. But now, although the typical pair-cycle has length a, and there are $(a - 2)/2$ choices for the gap between x and y, there is also the possibility that x and y are $a/2$ units apart in the cycle, so that the pair-cycle has length $a/2$. The contribution to $Z([S_n]^2)$ is

$$\prod_{a \text{ even}} x_a^{c(a)(a-2)/2} x_{a/2}^{c(a)}. \tag{2.48}$$

Combining the formulas (2.45), (2.45), (2.45), and (2.45), we obtain a formula for the cycle index of $[S_n]^2$:

$$Z([S_n]^2) = \frac{1}{n!}\sum_c h(c) \prod_{1\le a<b\le n} x_{\mathrm{lcm}(a,b)}^{c(a)c(b)\gcd(a,b)} \prod_{1\le a\le n} x_a^{a\binom{c(a)}{2}} \cdot \prod_{a \text{ odd}} x_a^{c(a)(a-1)/2} \prod_{a \text{ even}} x_a^{c(a)(a-2)/2} x_{a/2}^{c(a)}. \tag{2.49}$$

The cycle indexes for $2 \le n \le 4$ are easily calculated by hand:

$$Z([S_2]^2) = \frac{1}{2}(2x_1) \tag{2.50}$$

$$Z([S_3]^2) = \frac{1}{6}(x_1^3 + 3x_1x_2 + 2x_3) \tag{2.51}$$

$$Z([S_4]^2) = \frac{1}{24}(x_1^6 + 9x_1^2x_2^2 + 8x_3^2 + 6x_4x_2). \tag{2.52}$$

Let's examine the polynomial $Z([S_4]^2)$. The 24 permutations of the set {1, 2, 3, 4} induce permutations of the six unordered pairs of elements. The identity permutation results in six fixed pairs. The six permutations of type 2 + 1 + 1 result in two fixed points and two 2-cycles in the set unordered pairs. The three permutations of type 2 + 2 result in two fixed points and two 2-cycles. The eight permutations of type 3 + 1 result in two 3-cycles. The six 4-cycles result in one 2-cycle and one 4-cycle.

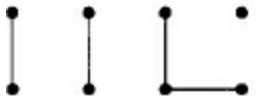
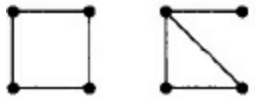
We apply Pólya's theorem to the cycle index $Z([S_4]^2)$ to enumerate nonisomorphic graphs of order 4 by number of edges:

$$Z([S_4]^2)[x_i \leftarrow y_1^i + y_2^i] = y_1^6 + y_1^5y_2 + 2y_1^4y_2^2 + 3y_1^3y_2^3 + 2y_1^2y_2^4 + y_1y_2^5 + y_2^6. \tag{2.53}$$

The coefficient of $y^k{}_1y^{6-k}{}_2$ in this enumerating polynomial is the number of nonisomorphic graphs of order 4 with k edges and $6 - k$ non-edges. The symmetry in the polynomial comes from the fact that G and H are isomorphic if and only if their complements $\overline{G}$ and $\overline{H}$ are isomorphic.

We verify formula (2.53) by presenting in Figure 2.6 all nonisomorphic graphs of order 4 arranged by number of edges.

Figure 2.6 Graphs of order 4 and their enumerating monomials.

graph	corresponding monomial
	y_2^6
	$y_1y_2^5$
	$2y_1^2y_2^4$
	$3y_1^3y_2^3$
	$2y_1^4y_2^2$
	$y_1^5y_2$
	y_1^6

The total number of nonisomorphic graphs of order 4 is

$$\begin{aligned} Z([S_4]^2)[x_i \leftarrow 2] &= Z([S_4]^2)[y_1 \leftarrow 1, y_2 \leftarrow 1] \\ &= 1+1+2+3+2+1+1 \\ &= 11. \end{aligned}$$

The number of nonisomorphic graphs. The number of nonisomorphic graphs of order n is

$$Z([S_n]^2)[x_i \leftarrow 2].$$

Let $g(n)$ be the number of nonisomorphic graphs on n vertices. Table 2.3 gives the value of $g(n)$ for small n. We will show that

Table 2.3 The number of nonisomorphic graphs.

order	number of graphs
1	1
2	2
3	4
4	11
5	34
6	156
7	1044

8	12346
9	274668
10	12005168

$$g(n) \sim \frac{2^{\binom{n}{2}}}{n!}.$$

Because $2^{\binom{n}{2}}$ is the number of labeled graphs, this means that almost all graphs have no nontrivial automorphisms. In fact, $g(n) \sim c(n)$, where $c(n)$ is the number of nonisomorphic connected graphs of order n. See [14].

Open problem. Find a formula for the number of nonisomorphic graphs of order n that contain a triangle.

Open problem. Find a formula for the number of nonisomorphic planar graphs of order n.

EXERCISE

2.63 Use formula (2.49) to find the number of graphs of order 5 (up to isomorphism).

2.64 How many ways are there to color the edges of K_6 with three colors (up to isomorphism)?

2.65 How many ways can the edges of the graph $K_{3,3}$ be colored with nine colors (up to isomorphism of the graph)?

2.66 Write the generating function for nonisomorphic multigraphs in terms of $Z(S_n)$. (A *multigraph* is a graph in which pairs of vertices can be joined by more than one edge.)

2.67 How many nonisomorphic multigraphs of order 4 have at most five edges?

2.68 Prove the identity

$$Z(S_n) = \frac{1}{n}\sum_{k=1}^{n} x_k Z(S_{n-k}).$$

2.69 (a) Let $Z(S_0) = 1$. Prove the identity $\sum_{n=0}^{\infty} Z(S_n) = \exp \sum_{k=1}^{\infty} x_k/k$.

(b) Let a_n be the number of onto functions $f : X \to X$, where $|X| = n$, with the property that $f(f(x)) = x$ for all $x \in X$. Show that the exponential generating function for the sequence $\{a_n\}$ is $\sum_{n=0}^{\infty} a_n x^n/n! = \exp(x + x^2/2)$.

Note that a_n is the number of involutions of an n-set and also the number of $n \times n$ symmetric permutation matrices. See Exercise 2.45.

2.9 Symmetries in domain and range

Suppose that X and Y are finite nonempty sets ($|X| = m$, $|Y| = n$) acted upon by finite groups G and H, respectively. We picture this situation with the following diagram:

$$f: \begin{array}{ccc} G & & H \\ \downarrow & & \downarrow \\ X & \longrightarrow & Y \end{array}$$

We want to define what it means for two functions to be the same under these group actions. To this end, we define a new action H^G on the set of functions $Y^X = \{f: X \to Y\}$ as follows: if $g \in G$, $h \in H$,

and $f \in Y^X$, then $(h^g f)(x) = hf(g^{-1}x)$. (The reader should check that the two axioms for a group action are satisfied.) We say that two functions are *equivalent* if they are in the same orbit of the action and *inequivalent* otherwise. This definition is a generalization of the labeled/unlabeled set paradigm. If G is the symmetric group S_m, then X is unlabeled, and if G is the identity group E_m, then X is labeled. Likewise, if H is S_n, then Y is unlabeled, and if H is E_n, then Y is labeled. For any groups G and H, Burnside's lemma gives the number of inequivalent functions:

$$N = \frac{1}{|G||H|} \sum_{g \in G} \sum_{h \in H} \Omega(g, h), \tag{2.54}$$

where $\Omega(g, h)$ is the number of functions fixed by h^g.

If f is fixed by h^g, then $f(g^{-1}x) = hf(x)$ for all $x \in X$. Thus $f(x) = y$ implies that $f(g^{-1}x) = hf(x) = hy$, which in turn implies that $f(g^{-2}x) = h^2 y$. In general, $f(g^{-i}x) = h^i y$. It follows that if x is in a cycle of length i in g and y is in a cycle of length j in h, then j divides i. Furthermore, we have shown that the correspondence between the two cycles is completely determined by the relation $f(x) = y$. There are j choices for the image of x. Suppose that g has cycle structure $(c(1),\ldots,c(m))$ and h has cycle structure $(d(1),\ldots,d(n))$. Then, given i and a particular cycle of length i in g, the number of fixed functions is

$$m_i(h) = \sum_{j|i} jd(j), \tag{2.55}$$

and the total number of fixed functions is

$$\Omega(g, h) = \prod_i m_i(h)^{c(i)}. \tag{2.56}$$

Formulas (2.54) through (2.56) combine to yield the following formula of N. G. de Bruijn (1918–2012).

De Bruijn's formula. If finite groups G and H act on finite nonempty sets X and Y, respectively, then the number of inequivalent functions is given by

$$N = \frac{1}{|H|} \sum_{h \in H} Z(G)[x_i \leftarrow m_i(h)]. \tag{2.57}$$

We apply de Bruijn's formula to the problem of counting self-symmetric graphs of order n, that is, graphs G for which G is isomorphic to its complement $\overline{G}$. In the previous section we determined the cycle index $Z([S_n]^2)$ and the number of nonisomorphic graphs of order n, i.e., $g(n) = Z([S_n])[x_i \leftarrow 2]$. If $[S_n]^2$ acts on the set $X = \{1,\ldots,n\}$ and S_2 acts on $Y = 11$, then functions correspond to nonisomorphic graphs where G and its complement $\overline{G}$ are regarded as the same. Hence, the number of such functions is

$$N(n) = \frac{1}{2} \sum_{h \in S_2} Z([S_n]^2)[x_i \leftarrow m_i(h)]. \tag{2.58}$$

Let $g^*(n)$ be the number of self-complementary graphs of order n. As $2N(n)$ counts nonisomorphic graphs with each self-complementary graph counted twice, we obtain the formula

$$g^*(n) = 2N(n) - g(n). \tag{2.59}$$

For example, $g^*(4) = 2N(4) - g(4) = 2 \cdot 6 - 11 = 1$, and the unique self-complementary graph of order 4 is the path P_4.

EXERCISE

2.70 How many self-complementary graphs are there of orders 5 and 6?

2.71 Use a computer to calculate the number of self-complementary graphs of orders 7 through 12.

2.10 Asymmetric graphs

Our formula for the number of nonisomorphic graphs of order n is

$$g(n) = \frac{1}{n!} \sum_c h(c) 2^{q(c)}, \tag{2.60}$$

where $c = (c_1,\ldots,c_n)$ is the cycle type of a permutation of $\{1,\ldots,n\}$, the number of permutations of such a cycle type is

$$h(c) = \frac{n!}{\prod_{k=1}^{n} k^{c_k} c_k!}, \tag{2.61}$$

and

$$q(c) = \sum_k \left\lfloor \frac{k}{2} \right\rfloor c_k + \sum_k k \binom{c_k}{2} + \sum_{r<s} \gcd(r,s) c_s c_s. \tag{2.62}$$

We will prove that

$$g(n) \sim \frac{2^{\binom{n}{2}}}{n!}. \tag{2.63}$$

This result means that the typical unlabeled graph has no nontrivial symmetries and hence is an *asymmetric graph*.

The lower bound is trivial:

$$\frac{2^{\binom{n}{2}}}{n!} \le g(n).$$

Now we prove the upper bound. Assume that the permutation has $n - j$ fixed points, where $0 \le j \le n$. The case $j = 0$ gives the term $2^{\binom{n}{2}}/n!$. We will show that the other terms are bounded above by expressions which, upon division by $2^{\binom{n}{2}}/n!$, tend to 0 as $n \to \infty$.

We have

$$q(c) \le \sum_k \left[\frac{k}{2}c(k) + \frac{k}{2}c(k)(c(k)-1)\right] + \sum_{r<s} \min(r,s)c(r)c(s)$$
$$\le \frac{1}{2}\sum_k kc(k)^2 + \sum_{r<s}\left(\frac{r+s}{2}\right)c(r)c(s)$$
$$= \frac{1}{2}\sum_k kc(k)\sum_k c(k)$$
$$= \frac{1}{2}n\sum_k c(k).$$

The number of permutations with $n-j$ fixed points is bounded above by

$$\binom{n}{n-j}j! = \frac{n!}{(n-j)!} \le n^j = 2^{j\log_2 n}.$$

The case $j = 2$ yields the exponent

$$q(c) = 1 + \binom{n-2}{2} + 1\cdot(n-2)\cdot 1 = \binom{n}{2} - n + 2.$$

Upon dividing by $2^{\binom{n}{2}}/n!$, the contribution is bounded by $2^{-n+2+2\log_2 n}$, which tends to 0 as $n \to \infty$.

Now let's look at the case $j \ge 3$, so that $1 - j/2 < 0$. We obtain

$$\sum_k c(k) \le n - j + \frac{j}{2} = n - \frac{j}{2}$$

and

$$q(c) \le \frac{1}{2}n\left(n - \frac{j}{2}\right) = \binom{n}{2} + \frac{1}{2}n\left(1 - \frac{j}{2}\right).$$

Upon dividing by $2^{\binom{n}{2}}/n!$, the contribution is bounded by $2^{\frac{1}{2}n(1-j/2)+j\log_2 n}$, which tends to 0 as $n \to \infty$.

Erdős and Rényi theorem (1963). Almost all unlabeled graphs are asymmetric:

$$g(n) \sim \frac{2^{\binom{n}{2}}}{n!}.$$

EXERCISE

2.72 Find an asymmetric graph with six vertices. Show that no graph with five vertices is asymmetric.

2.73 Use a computer to compare $g(n)$ and $2^{\binom{n}{2}}/n!$ for $2 \le n \le 20$.

2.74 Estimate $g(100)$.

Notes

The notation for falling factorial and rising factorial varies considerably. The falling factorial

function is often written $(x)_n$.

James Sylvester (1814–1897) introduced the notion of Ferrers diagrams in 1853. Apparently they were discovered by Norman Ferrers (1829–1903).

G. H. Hardy and Srinivasa Ramanujan (1887–1920) were the first mathematicians to find an explicit formula for $p(n)$. The most elementary asymptotic estimate is $\log_e p(n) \sim \pi(2n/3)^{1/2}$. See [12] for details.

In 1937 George Pólya (1887–1985) published the formula for enumerating graphs in connection with a problem concerning the number of chemical isomers. The language of his counting theory was quite descriptive. A function from X to Y is called a *configuration*. The elements of X are *places* and the elements of Y are *figures*. The *store inventory* is $\sum_{j=1}^{n} y_j^i$ and the *pattern inventory* is $Z(G)[x_i \leftarrow \sum_{j=1}^{n} y_j^i]$. Thus the basic problem is to find the number of G-inequivalent patterns. An alternative approach to the Pólya theory of counting was undertaken independently by John H. Redfield (1879–1944) in 1927.

CHAPTER 3

THE PIGEONHOLE PRINCIPLE

The pigeonhole principle is an important key to solving many problems in combinatorics. In this chapter, we discuss several versions of the pigeonhole principle and give many applications.

3.1 The principle

The most important theorem in existential combinatorics is also the simplest: the pigeonhole principle. It occurs in many variations, a few of which we discuss here, and says that not every element in a set is below average and not every element is above average. We now state and prove a general version.

Pigeonhole principle. If $f : A \to B$ is a function, with A and B finite nonempty sets, then the following two statements hold:

(1) 'There exists $b \in B$ with $|f^{-1}(b)| \geq |A|/|B|$.

(2) There exists $b \in B$ with $|f^{-1}(b)| \leq |A|/|B|$.

Proof. We prove (1) by contradiction. Suppose that $|f^{-1}(b)| < |A|/|B|$ for all $b \in B$. Then

$$|A| = \sum_{b \in B} |f^{-1}(b)| < \frac{|A|}{|B|} \cdot |B| = |A|.$$

We conclude that $|A| < |A|$, an absurdity. Therefore, our assumption that $|f^{-1}(b)| < |A|/|B|$ for all $b \in B$ is false. Hence, there exists $b \in B$ with $|f^{-1}(b)| \geq |A|/|B|$ for all.

We prove (2) by replacing "<" by ">" in the above argument.

EXAMPLE 3.1 Sets of cards

A popular board game features cards of three suits: cannon, horse, and soldier. A "set" consists of three horses, three soldiers, three cannons, or one card of each suit. It is possible to have four cards without possessing a set, e.g., two horses and two soldiers. Prove that any five cards contain a set.

Solution: Let the three suits be designated by *C, H*, and *S*. If the five cards do not include one card of each suit, then at least one suit is absent, say *S*. Therefore, we may define a function f: $\{a, b, c, d, e\} \to \{C, H\}$ from the five cards to their respective suits. (The function isn't necessarily onto.) By the pigeonhole principle, the preimage of one suit contains at least three cards. These cards constitute a set.

Nonuniform pigeonhole principle. If f: $A \to B$ is a function from a finite nonempty set A to an n-set $B = \{b_1, b_n\}$, then the following two statements hold:

(1) If $|A| = (\alpha_1 - 1) + \cdots + (\alpha_n - 1) + 1$, then $|f^{-1}(b_i)| \geq \alpha_i$ for some i.

(2) If $|A| = (\alpha_1 + 1) + \cdots + (\alpha_n - 1) - 1$, then $|f^{-1}(b_i)| \leq \alpha_i$ for some i.

Proof. (1) If the inequality holds for no i, then

$$|A| = \sum_{i=1}^{n} |f^{-1}(b_i)| \leq \sum_{i=1}^{n} (\alpha_i - 1) < |A|,$$

a contradiction.

(2) is proved similarly.

If $|A| = |B| + 1$, then the following special case of the pigeonhole principle results.

Pigeonhole principle (special case). If f: $A \leftarrow B$ is a function and $|A| = |B| + 1$, then there exists $b \in B$ with $|f^{-1}(b)| \geq 2$. In other words, $f(a_1) = f(a_2)$ for some distinct $a_1, a_2 \in A$.

This version of the pigeonhole principle is often paraphrased as: "If $n + 1$ objects are placed in n pigeonholes, then at least one pigeonhole must contain at least two objects." Hence the term *pigeonhole principle*.

We now give a pigeonhole principle proof of a very old but interesting result in number theory. For different proofs, see [21].

Approximation theorem. For any real number α and $n \in \mathbf{N}$, there exist integers p and q with $1 \leq q \leq n$ and $|\alpha - p/g| < 1/qn \leq 1/q^2$.

Proof. Define $f: \mathbf{N}_{n+1} \to \{[0, \frac{1}{n}), [\frac{1}{n}, \frac{2}{n}), \ldots, [\frac{n-1}{n}, 1)\}$ by letting $f(j)$ be the subinterval of $[0, 1)$ which contains $j\alpha$ — $[j\alpha]$. (Note that $[x]$ is the greatest integer less than or equal to x.) The pigeonhole principle implies the existence of j, $k(j > k)$ with $f(j) = f(k)$. According to the definition of the function f, there exists a positive integer p with $|j\alpha - k\alpha - p| < 1/n$. Letting $q = j - k$ (so that $1 \leq q \leq n$), the inequality becomes $|\alpha - p/q| < 1/qn \leq 1/q^2$.

This theorem is used to ensure good rational approximations to irrational numbers α. For example, taking $\alpha = \pi$ and $n = 10$, the theorem guarantees the existence of a rational number p/q with $|\pi - p/q| < 1/10q \leq 1/q^2$. In fact, the well-known approximation 22/7 satisfies the inequality. Can you find an approximation to $\alpha = \sqrt{2}$ with $n = 10$?

In the preceding versions of the pigeonhole principle we have assumed that both the domain and the codomain are finite sets. If the domain is infinite, then the following highly useful infinitary pigeonhole principle results.

Infinitary pigeonhole principle. If f: $A \to B$ is a function from an infinite set A to a finite set B, then there exists $b \in B$ with $f^{-1}(b)$ infinite.

EXERCISES

3.1 Using the pigeonhole principle, show that some positive integral power of 17 ends in 0001 (base 10).

3.2 Let q be an odd integer greater than 1. Show that there is a positive integer n such that $2^n - 1$ is a multiple of q.

3.3 Let $A_1,\ldots,A_{100}$ be subsets of a finite set S, each with $|A_i| > \frac{2}{3}|S|$. Prove that there exists $x \in S$ with x contained in at least 67 of the A_i. Show that 67 is the best possible result.

3.4 Prove that if S is a subset of $\{1,\ldots,2n\}$ and $|S| > n$, then there exist $x, y \in S$ with x and y relatively prime.

3.5 Prove that if S is a subset of $\{1,\ldots,2n\}$ and $|S| > n$, then there exist $x, y \in S$ with x a divisor of y.

3.6 (Putnam Competition, 1964) Let S be a set of $n > 0$ elements, and let $A_1, A_2,\ldots,A_k$ be a family of distinct subsets, with the property that any two of these subsets meet. Assume that no other subset of S meets all of the A_i. Prove that $k = 2^{n-1}$.

This result is generalized in many interesting ways in [3].

3.7 Prove the infinitary pigeonhole principle.

Hint: Assume that $f^{-1}(b)$ is finite for all $b \in B$ and obtain a contradiction.

3.8 Prove that every polyhedron has two faces with the same number of edges.

3.9 Let A be an $m \times n$ matrix with distinct real number entries in increasing order in each row from left to right. Rearrange the elements of each column of A so that they are in increasing order from top to bottom; call the resulting matrix A'. Show that the elements of each row of A' are in increasing order from left to right.

3.10 An $n \times n$ binary matrix contains a 1 in every row, column, and diagonal (diagonals of every length are considered here). What is the minimum number of 1′s in this matrix?

3.11 Let L_1 be a two-row array of positive integers

$$\begin{array}{cccc} a_1 & a_2 & \ldots & a_m \\ b_1 & b_2 & \ldots & b_m \end{array}$$

where the a_i are distinct integers written in increasing order. Let $c_1,\ldots,c_n$ ($c_1 \le c_2 \le \cdots c_n$) be the list of all integers that occur in L_1, and for $1 \le i \le m$, let d_i be the number of occurrences of c_i in L_1. Let L_2 be the array

$$\begin{array}{cccc} c_1 & c_2 & \ldots & c_n \\ d_1 & d_2 & \ldots & d_n. \end{array}$$

For example, if L_1 is the array

$$\begin{array}{ccccccc} 1 & 2 & 5 & 6 & 8 & 10 & 11 \\ 3 & 3 & 1 & 4 & 1 & 3 & 6 \end{array}$$

then L_2 is the array

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 8 & 10 & 11 \\ 3 & 1 & 3 & 1 & 1 & 2 & 1 & 1 & 1. \end{array}$$

Starting with any array L_1, the array L_2 is created as indicated above. Then the operation is

repeated on L_2 to form a new array L_3, and so on. Show that the number of distinct arrays produced in this manner is always finite. We say that each sequence of arrays eventually "goes into a loop." Show that a loop always consists of one, two, or three arrays.

3.2 The lattice point problem and SET®

In this section, we apply the pigeonhole principle to problems concerning lattice points in Euclidean space.

A *lattice point* in the plane is an ordered pair $p = (x, y)$ with integer coordinates x and y.

EXAMPLE 3.2 A lattice point midpoint

Let p_1, p_2, p_3, p_4, p_5 be five lattice points in the plane. Prove that the midpoint of the line segment p_ip_i determined by some two distinct lattice points p_i and p_i is also a lattice point.

Solution: Define a function

$$f: \{p_1, p_2, p_3, p_4, p_5\} \longrightarrow \{(0,0), (0,1), (1,0), (1,1)\}$$

by mapping p_i to the ordered pair (x_i mod 2, y_i mod 2). By the pigeonhole principle, some two points p_i and p_i have the same image. These points satisfy the requirement of the problem, for the midpoint of p_ip_j is $((x_i + x_j)/2, (y_i + y_j)/2)$. Both coordinates are integers because x_i, x_j have the same parity and y_i, y_j have the same parity.

The "five" in the problem above is best possible in the sense that one can find four lattice points determining no lattice midpoint, e.g., (0, 0), (0, 1), (1, 0), (1, 1).

EXAMPLE 3.3 A lattice point centroid

The *centroid* of three points $p_i = (x_i, y_i)$, $pj = (x_i, y_i)$, $p_k = (x_k, y_k)$ is $((x_i + x_j + x_k)/3, (y_i + y_i + y_k)/3)$. What is the minimum number n of lattice points, some three of which must determine a lattice point centroid?

Solution: We show that $n \leq 13$ by defining a function f: $\{p_1,\ldots,p_{13}\} \rightarrow \{0, 1, 2\}$ which maps p_i to the residue class modulo 3 of its first coordinate. By the pigeonhole principle, some five lattice points, say p_1, p_2, p_3, p_4, p_5, have the same image. By the analysis of the suits in the board game mentioned in Example 3.1, some three of these points, say p_1, p_2, p_3, have second coordinate residues 000, 111, 222, or 012. These three lattice points determine a lattice point centroid. Therefore, n exists and satisfies $n \leq 13$.

The determination of the minimum n that forces the existence of three points determining a lattice point centroid is a more difficult matter. It turns out that the minimum value is $n = 9$. The argument is carried out modulo 3. Thus, there are just nine possible ordered pairs from which to choose. The following list of ordered pairs (modulo 3) shows that $n > 8$:

$$(0,0), (0,0), (1,0), (1,0), (0,1), (0,1), (1,1), (1,1).$$

In order to prove $n \leq 9$, we must show that any nine points include three whose first coordinates and

second coordinates are of the form 000, 111, 222, or 012. The nine possible ordered pairs are conveniently represented by the nine non-ideal points of the order 3 projective plane of Figure 6.1. The 12 lines of the figure (excluding the line at infinity) correspond to triples of points which determine a lattice point centroid. If any ordered pair is chosen three times, these three ordered pairs determine a lattice point centroid. Therefore, let us assume that each ordered pair is chosen at most twice, and hence at least five different ordered pairs are chosen. By shuffling the rows and columns of the figure (if necessary), we may assume that three of the points are (0, 0), (1, 0), and (1, 1). If no three points are collinear, then we may not choose the points (2, 0), (0, 2), or (2, 2). This means that we must choose two of the three points (1, 1), (2, 1), and (1, 2). But any of these choices gives three collinear points: (0, 1), (1, 1), and (2, 1); (1, 0), (1, 1), and (2, 1); or (0, 0), (1, 2), and (2, 1).

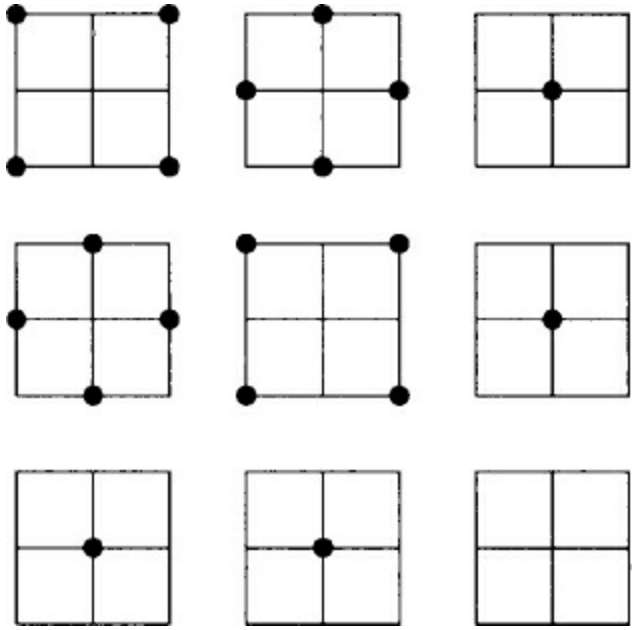
The d-dimensional generalization of the above problem calls for the minimum n such that, given any n lattice points in $\mathbf{R}^d$ (ordered d-tuples of integers), some k determine a lattice point centroid. Let $n(k, d)$ be the minimum such n. The existence of $n(k, d)$ is guaranteed by the pigeonhole principle, and, in fact, the pigeonhole principle yields an upper bound for $n(k, d)$ on the order of k^{d+1}. The set of d-tuples each of whose coordinates is 0 or 1, taken with multiplicity $k - 1$, establishes the lower bound $n(k, d) > (k - 1)2^d$.

Open problem. Find a formula for $n(k, d)$ for all k, d.

This question is known as the lattice point problem. Trivially, $n(1, d) = 1$. In 1961 Paul Erdős, A. Ginzburg, and A. Ziv proved that $n(k,1) = 2k - 1$. It has since been shown that $n(3, 3) = 19$; she Problem 6298, *American Mathematical Monthly* **89** (1982) 279–280. In 2003 Christian Reiher and Carlos di Fiore proved that $n(k, 2) = 4k - 3$. An exercise calls for a proof that $n(2, d) = 2^d + 1$.

A situation similar to the lattice point problem comes from a card game called SET®, created by the population geneticist Marsha Falco. The game is played with cards identified by four attributes: number, shape, color, and shading. Each attribute occurs with three possible values (e.g., shape = oval, diamond, or squiggle). Hence there are 81 cards. Thinking of the attributes as coordinates and the values as residue classes, the cards are represented as 4-tuples over $\mathbf{Z}_3$, e.g., (0, 1, 0, 1), (2, 1, 0, 2), and (1, 1, 1, 1). *A* "set" consists of three cards that, with respect to each attribute, all agree or all disagree. For example, the cards corresponding to (0, 1, 1, 2), (0, 2, 0, 2), and (0, 0, 2, 2) constitute a set. The game is similar to the lattice point problem with $k = 3$ and $d = 4$, but in the game no 4-tuple of coordinates is repeated. One can define a generalized version of the game in which there are d attributes each occurring with k possible values. Hence, there are k^d cards. A set consists of k cards that all agree or all disagree on each attribute. Equivalently, a card is a vector with d coordinates, each of which can take k values. A set is a collection of k vectors, which in each coordinate all agree or all disagree. Then we can define the generalized SET® function $n'(k, d)$ to be the minimum number of cards (or vectors) that guarantee a set. Little is known about the function $n'(k, d)$. The values $n'(3, 4) = 21$, $n'(3, 5) = 46$, and $n'(3, 6) = 113$ are known. A collection of 20 cards containing no set is shown in Figure 3.1 (the four-dimensional space is represented by a three-by-three array of three-by-three arrays). Information on SET may be found at http://www.setgame.com.

Figure 3.1 Twenty cards containing no set.



The functions $n(k, d)$ and $n'(k,d)$ are similar, although there are two differences. A minor difference is that in the card game each d-tuple occurs exactly once while in the lattice point problem repetition is allowed. A more important difference is that in the lattice point problem we require that a sum of lattice points be zero modulo k, while in the card game we require that the cards all agree or all disagree in each coordinate. For example, the lattice points (1, 0), (1, 1), (3, 0), and (3, 3) have a lattice point centroid, but the corresponding cards do not constitute a set (with $k = 4$). If $k = 3$, this difference disappears.

We mentioned that the value $n(3, 3) = 19$ has been proved. The corresponding fact in the card game is that $n'(3, 3) = 10$.

In the card game setting, there are 27 possible cards (each defined by three attributes that occur with three possible values). We would like to show that every collection of 10 cards contains a set. To translate back to the lattice point problem, assume that we have 19 ordered triples of numbers modulo 3. If no triple is repeated three times, then, by the pigeonhole principle, there must be at least 10 distinct triples. An exercise calls for a construction to show that $n(3, 3) > 18$ (and incidentally $n'(3, 3) > 9$).

It is convenient to represent the cards both as integers between 0 and 26 (inclusive) and as 3-tuples of elements in $\mathbf{Z}_3$. We define these 3-tuples as the base 3 digits of the numbers.

The number of 10-subsets of a 27-element set is $\binom{27}{10} = 8{,}436{,}285$, and this is too many subsets to test. We will discuss a way to reduce the number of 10-subsets which must be examined.

Without loss of generality, we may assume that $0 \equiv (0, 0, 0)$ is an element of each 10-subset. Also, we note that in any 10-subset of $\mathbf{Z}^3_3$ there exist three linearly independent vectors (a plane contains only nine points). By a linear transformation (a 3×3 matrix), we may map these three vectors to the specific vectors $1 \equiv (0, 0, 1)$, $3 \equiv (0, 1, 0)$, and $9 \equiv (1, 0, 0)$. Therefore, since linear transformations map sets to sets, we need only consider 10-subsets that contain these three points. Furthermore, each pair of the four points 0, 1, 3, 9 determines a third point which forms a set with that pair. There are $\binom{4}{2} \equiv 6$ such points, namely, $2 \equiv (0, 0, 1)$, $6 \equiv (0, 2, 0)$, $18 \equiv (2, 0, 0)$, $8 \equiv (0, 2, 2)$, $20 \equiv (2, 0, 2)$, and $24 \equiv (2, 2, 0)$. If any of these points is present in a 10-subset, then the 10-subset contains a set and we do not need to examine it. We can therefore exclude these six points from consideration.

To summarize, of the integers 0 through 26 (inclusive), the integers 0, 1, 3, and 9 are included in every 10-subset, and the integers 2, 6, 8, 18, 20, and 24 are excluded in every 10-subset. Hence, we now have only $\binom{27-4-6}{10-4} = \binom{17}{6} = 12{,}376$ subsets to examine.

A computer run can check these subsets and thereby show that every 10-subset contains a set, and therefore $n'(3, 3) = 10$ and $n(3, 3) = 19$.

Open problem. Find $n'(4, 3)$.

EXERCISES

3.12 (Putnam Competition, 1971) Let there be given nine lattice points (points with integer coordinates) in three-dimensional Euclidean space. Show that there is a lattice point on the interior of one of the line segments joining two of these points.

3.13 Prove that $n(2, d) = 2^d + 1$.

3.14 Prove that $n(3, 3) > 18$. In fact, $n(3, 3) = 19$.

3.15 Prove that $(k-1)2^d + 1 \leq n(k,d) < (k-1)k^d + 1$

3.16 Prove that if k is a power of 2, then $n(k, d) = (k-1)2^d + 1$.

3.17 Prove that $n(3, d) = 2n'(3, d) - 1$.

3.18 Prove that $n^1(3, d+1) - 1 \geq 2(n'(3, d) - 1)$.

3.19 Prove that $(k-1)^d < n'(k, d) < k^d$.

3.20 Prove that $n'(k, 2) = (k-1)^2 + 1$.

3.21 Prove that given any $n^d - n^{d-1} + 1$ d-tuples from the set $S = \{1, 2,\ldots,n\}$, there exist n which, in each coordinate, are a permutation of S. Show that the result is not true for $n^d - n^{d-1}$ d-tuples.

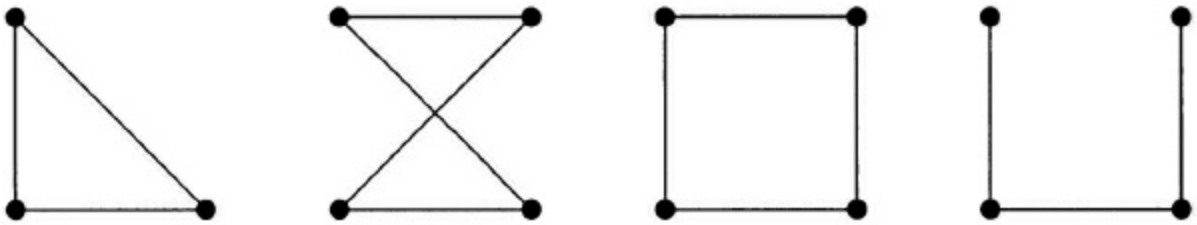
3.3 Graphs

Graph theory began in 1736 when Leonhard Euler (1707–1783) solved the famous concerning a certain system of seven bridges over the river Pregel. See [16]. In the last 50 years there has been an explosion in graph theory research and applications. Today, there are many areas of graph theory research including algebraic graph theory, extremal graph theory, and topological graph theory. Within combinatorics, graph theory is closely related to design theory, Ramsey theory, and coding theory. In this section we give some basic definitions and an indication of the deeper results of graph theory which will be studied in the next chapter.

A *graph* G is an ordered pair (V, E), consisting of a *vertex set* V and an *edge set* $E \subseteq [V]^2$. Vertices are also called *points* or *nodes*. Edges are also called *lines* or *arcs*. In our definition of graph there are no loops or multiple edges. In a drawing of a graph, two vertices x and y are joined by a line if and only if $\{x, y\} \in E$. Two vertices joined by a line are said to be *adjacent*; if they are not joined by a line, they are *nonadjacent*. If $|V| = p$ and $|E| = q$, then we say that G has *order* p and *size* q. The *degree* of a vertex v, denoted $\delta(v)$, is the number of edges incident to v. The *complement* $\overline{G}$ of a graph G has the same vertex set as G, but two vertices are adjacent in $\overline{G}$ if and only if they are nonadjacent in G.

Certain graphs occur so frequently that they require names. The *complete graph* K_n, consists of n vertices and all $\binom{n}{2}$ possible edges. The *complete bipartite graph* $K_{m,\,n}$ consists of a set A of m points, a set B of n points, and all the mn edges between A and B. The *infinite complete graph* K_∞ contains a countably infinite set of points and all possible edges. Likewise, the *infinite complete bipartite graph* $K_{\infty,\,\infty}$ contains countably infinite sets A and B and all edges between A and B. The *cycle* C_n consists of n vertices connected by n edges in a cyclical fashion. The *path* P_n is C_n minus an edge. Figure 3.2 illustrates some of these graphs. For general references on graph theory, see [13], [5], and [28].

Figure 3.2 A complete graph, a complete bipartite graph, a cycle, and a path.



One of the most elementary propositions of graph theory is called the "Handshake Theorem." If some people in a group shake hands, then there will be two people who shake the same number of hands.

Handshake Theorem. In any graph G with a finite number of vertices, some two vertices have the same degree.

Proof. Suppose that G has p vertices. Then each vertex has degree equal to one of the numbers $0,\ldots,p-1$. However, it is impossible for G to have both a vertex of degree 0 and a vertex of degree $p-1$. Therefore the list of degrees of the p vertices contains at most $p-1$ different numbers. By the pigeonhole principle, some two vertices have the same degree.

If $\delta(v)$ is the same for all vertices, then we say that G is *regular* of *degree* $\delta(v)$. Note that complete graphs and cycles are regular.

If G is any finite graph, the *independence number* $\alpha(G)$ is the maximum possible number of pairwise nonadjacent vertices of G. The *chromatic number* $\chi(G)$ is the minimum number of colors in a coloring of the vertices of G with the property that no two adjacent vertices share the same color.

Here is another simple theorem proved by the pigeonhole principle.

Theorem. In any graph G with p vertices, $p \leq \alpha(G)\chi(G)$.

Proof. Consider the vertices of G as partitioned into $\chi(G)$ color classes. By the pigeonhole principle, one of the classes must contain at least $p/\chi(G)$ vertices, and these vertices are pairwise nonadjacent. Thus $\alpha(G) \geq p/\chi(G)$ and the result follows immediately.

Equality in the above theorem holds, for example, when G consists of the vertices and edges of a cube.

The famous "four color theorem," proved in 1976 by Kenneth Appel and Wolfgang Haken, is the statement that $\chi(G) \leq 4$ for any planar graph G. (A graph is *planar* if it can be drawn in the plane without edge crossings.) Combining this result with the theorem on the independence number and chromatic number of a graph, we arrive at the relation $\alpha(G) \geq p/4$ for any planar graph G. We can turn a planar graph into a planar map by placing a territory at each vertex and allowing two territories to share a common boundary when the two vertices in the graph are adjacent. In terms of maps, Appel and Haken's result is that every map can be colored with four colors so that no two bordering territories have the same color. It follows from the theorem on the independence number and chromatic number of a graph that any planar map on p vertices contains at least $p/4$ territories no two of which share a border.

A *path* in a graph G from vertex v_0 to vertex v_n is a sequence of distinct edges

$$\{v_0, v_1\}, \{v_1, v_2\}, \ldots, \{v_{n-1}, v_n\}.$$

The path is *simple* if the vertices v_1, v_2,…,v_n are distinct. *A circuit* is a path from v to v for some vertex v. A *simple* circuit is a cycle. We say that G is *connected* if there is a path between every two vertices. Note that each of the graphs in Figure 3.2 is connected.

The next result, a special case of Turán's theorem called Mantel's theorem, foreshadows Ramsey's theorem of the next chapter.

We consider an *extremal property* of graphs. How many edges are possible in a triangle-free graph G on $2n$ vertices? Certainly, G can have n^2 edges without containing a triangle: let G be the complete bipartite graph $K_{n,n}$, consisting of two sets of n vertices each and all the edges between the two sets. Indeed, n^2 turns out to be the maximum possible number of edges. That is, if G has $n^2 + 1$ edges, then G contains a triangle. This we prove by mathematical induction using the pigeonhole principle.

Mantel's theorem (1907). If a graph G of order $2n$ has $n^2 + 1$ edges, then G contains a triangle.

Proof If $n = 1$, then G cannot have $n^2 + 1$ edges; hence the statement is vacuously true. Assuming the result for n, we now consider a graph G on $2(n + 1)$ vertices with $(n + 1)^2 + 1$ edges. Let x and y be adjacent vertices in G, and let H be the restriction of G to the other $2n$ vertices. If H has more than n^2 edges, then we are finished by the induction hypothesis. Suppose that H has at most n^2 edges, and therefore at least $2n + 1$ edges join x and y to vertices in H. By the pigeonhole principle, there exists a vertex z in H that is adjacent to both x and y. Hence G contains the triangle xyz.

Theorem. Up to isomorphism, $K_{n,n}$, is the only triangle-free graph with $2n$ vertices and n^2 edges.

Proof. The argument uses induction on n and the previous theorem. For $n = 1$, the result is trivially true. Assume the result holds for n. Let G be a graph on $2(n + 1)$ vertices with $(n + 1)^2$ edges and no triangle. Let u and v be connected vertices in G. Let H be the graph restricted to the other $2n$ vertices. By the previous theorem, H has at most n^2 edges. However, if H has less than n^2 edges, then there are more than $2n$ edges between the set $\{u, v\}$ and H; by the pigeonhole principles, there exists a triangle. Hence, H has exactly n^2 edges, H is isomorphic to $K_{n,n}$, and there are exactly $2n$ edges from $\{u, v\}$ to

H. The reader can now show that u and v are each joined by n edges to H, with u joined to one independent set in H and v joined to the other. Therefore, G is isomorphic to $K_{n+1,n+1}$. This completes the induction and the proof.

EXERCISES

3.22 Show that the theorem on the independence number and chromatic number of a graph does not hold for infinite graphs.

3.23 Find two nonisomorphic graphs with $p = 12$, $\alpha(G) = 3$, and $\chi(G) = 4$. This shows that the upper bound of the theorem on the independence number and chromatic number of a graph may be met by nonisomorphic graphs.

3.24 Use the infinitary pigeonhole principle to prove that if G is a countably infinite graph, then at least one of $\alpha(G)$ and $\chi(G)$ must be infinite.

3.25 Prove that if G is a graph with $\delta(v) \geq p/2$ for every vertex v, then G is connected.

3.26 (G. A. Dirac, 1952) Show that under the hypothesis of the previous exercise, G contains a subgraph isomorphic to C_p. Such a subgraph is called a *Hamiltonian circuit*, after the mathematician William Rowan Hamilton (1805–1865).

3.27 A *tree* is a connected graph with no cycles. Prove that in a tree with p vertices and q edges, $p = q + 1$.

3.28 A graph is called "cubic" if every vertex is of degree 3. Prove that the edges of any cubic Hamiltonian graph (one with a Hamiltonian circuit) can be colored with three colors so that no two edges of the same color have a common vertex.

3.4 Colorings of the plane

The concept of this section, partitions of the plane, foreshadows Van der Waerden's theorem of Chapter 4.

Suppose that the plane is partitioned into two (disjoint) subsets G and R (green and red). We will show that one of the two subsets contains the vertices of a Euclidean rectangle with sides parallel to the coordinate axes. In fact, partitioning the whole plane is unnecessary. The same result follows if we partition just the 21 lattice points of $\mathbf{N}_7 \times \mathbf{N}_3$ into two subsets, so let us assume only that. We say that each lattice point is "colored" either G or R.

Theorem. If the 21 lattice points of $\mathbf{N}_7 \times \mathbf{N}_3$ are colored G and R, there exist four points, all the same color, lying on the vertices of a rectangle with sides parallel to the coordinate axes.

Proof. Each column of three points in this lattice contains three points of color G, two G's and one R, two R's and one G, or three R's. For the moment, the relevant fact is that there is a majority of G's or a majority of R's in each column. Let us refer to a column as a G-majority or an R-majority column. By the pigeonhole principle, some four columns are G-majority or some four columns are R-majority. Without loss of generality, suppose that there are four G-majority columns. We will show that there are four points all colored G which are the vertices of a rectangle. If any of the four G-majority

columns contain three points colored *G*, then we handicap ourselves by changing the color of an arbitrary point to *R*. (Our result will follow even with this handicap.) Now we have four columns which each contain exactly two points colored *G* and one point colored *R*. There are three possible patterns for the configuration of the points: *GGR, GRG*, and *RGG*. By the pigeonhole principle, there are two columns with the same color pattern. The four *G*'s in these columns are the vertices of a rectangle with sides parallel to the coordinate axes.

It is easy to see that neither the lattice $\mathbf{N}_7 \times \mathbf{N}_2$ nor the lattice $\mathbf{N}_6 \times \mathbf{N}_3$ is sufficient, when 2-colored, to force the existence of four monochromatic points on the vertices of a rectangle with sides parallel to the coordinate axe's. Thus, the lattice $\mathbf{N}_7 \times \mathbf{N}_3$ is *minimal* with respect to this property.

EXERCISES

3.29 Exhibit 2-colorings that show that neither the lattice $\mathbf{N}_7 \times \mathbf{N}_2$ nor the lattice $\mathbf{N}_6 \times \mathbf{N}_3$ is sufficient to force the existence of four monochromatic points on the vertices of a rectangle with sides parallel to the coordinate axes.

3.30 Investigate the same question as in the previous exercise, with three colors instead of two.

3.31 (R. Bacher and S. Eliahou, 2009) Prove that no matter how $\mathbf{N}_{14} \times \mathbf{N}_{15}$ is 2-colored, there exist positive integers *i, j, k* such that the set

$$\{(i, j), (i + k, j), (i, j + k), (i + k, j + k)\}$$

is monochromatic. There exists a square with horizontal and vertical sides and monochromatic vertices.

Show that there is a 2-coloring of $\mathbf{N}_{14} \times \mathbf{N}_{14}$ that avoids a monochromatic square.

3.5 Sequences and partial orders

Every sequence of 10 *distinct integers contains an increasing subsequence of four integers or a decreasing subsequence of four integers* (*or both*). For example, the sequence 5, 8, –1, 0, 2, –4, –2, 1, 7, 6 contains the increasing subsequence –1, 0, 2, 7.

This proposition is an existence result. No matter which 10 integers are chosen, and no matter in what order they occur, there exists a specific type of subsequence, namely, a monotonic subsequence of four integers.

In this section, we apply the pigeonhole principle to two types of mathematical structures: sequences and partial orders. Our goal is to show that arbitrary sequences and partial orders contain highly nonrandom substructures. This is indicative of a basic principle of existential combinatorics: complete disorder is impossible.

A sequence (finite or infinite) is *increasing* (or *strictly increasing*) if $a_i < a_i$ for all $i < j$; *decreasing* (or *strictly decreasing*) if $a_j > a_j$ for all $i < j$; *monotonically increasing* if $a_i \leq a_j$ for all $i < j$; *monotonically decreasing* if $a_i \geq a_j$ for all $i < j$; and *monotonic* if it is either monotonically increasing or monotonically decreasing.

EXAMPLE 3.4

The sequence {1, 1, 0, 0, –1, –1,…} is monotonically decreasing. The sequence {1, 4, 9, 16, 25, 36,…} is strictly increasing.

We say that $\{b_1,\ldots,b_m\}$ is a *subsequence* of $\{a_1,\ldots,a_n\}$ if there exists a strictly increasing function f : $\{1,\ldots,m\} \to \{1,\ldots,n\}$ for which $b_i = a_{f(i)}$ for all i.

EXAMPLE 3.5

The sequence {1, 2, 3, 2} is a subsequence of the sequence {1, 4, 2, 3, 5, 2}.

Erdős–Szekeres theorem (1935). Let $m, n \in N$. Every sequence of $mn + 1$ real numbers contains a monotonically increasing subsequence of $m + 1$ terms or a monotonically decreasing subsequence of $n + 1$ terms (or both).

Proof. Suppose that $S = \{a_1,\ldots,a_{mn+1}\}$ is a sequence of real numbers. For $1 \le k \le mn + 1$, let i_k be the length of a longest monotonically increasing subsequence starting with a_k, and let d_k be the length of a longest monotonically decreasing subsequence starting with a_k. Then the ordered pairs (i_k, d_k) are distinct. For if $j < k$ and $a_j \le a_k$, then $i_j > i_k$, while if $j < k$ and $a_j \ge a_k$, then $d_j > d_k$. But by the pigeonhole principle, if $1 \le i_k \le m$ and $1 \le d_k \le n$, for all k, then some ordered pairs (i_k, d_k) are not distinct. Therefore, $i_k \ge m + 1$ or $d_k \ge n + 1$ for some k.

EXAMPLE 3.6

Taking $m = 3$ and $n = 3$, the Erdős—Szekeres theorem guarantees that a sequence S of $mn + 1 = 10$ real numbers contains a monotonic subsequence of four terms. If the 10 terms of S are distinct, then of course the subsequence will be strictly increasing or strictly decreasing. Thus, the sequence 5, 8, –1, 0, 2, –4, 1, 7, 6 contains the strictly increasing subsequence –1, 0, 2, 7.

The expression $mn + 1$ in the Erdős—Szekeres theorem is best possible in the sense that there exists a sequence of mn real numbers which contains neither a monotonically increasing subsequence of $m + 1$ terms or a monotonically decreasing subsequence of $n + 1$ terms. We form such a sequence by concatenating n sequences of m increasing terms in the following manner. For each $j \in \mathbf{N}_n$, let $S_j = \{a_{1j},\ldots,a_{mj}\}$ be an increasing sequence of m real numbers, and suppose that every term of S_j is greater than every term of S_k whenever $j < k$. Then the sequence

$$S = \{a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn}\}$$

contains no increasing subsequence of length $m + 1$ and no decreasing subsequence of length $n + 1$. In general, there are many sequences which avoid monotonic subsequences of these lengths. The question of the number of such sequences is answered by the theory of Young tableaux. See Notes.

Erdős–Szekeres theorem (infinitary version). Every infinite sequence of real numbers contains an infinite monotonic subsequence.

Proof. Let $S = \{a_1, a_2,\ldots\}$ be an arbitrary infinite sequence of real numbers. We will inductively

define an infinite monotonic subsequence of S. Relabel the elements of S as $\{a_{11}, a_{12},\ldots\}$. By the infinitary pigeonhole principle, there exists an infinite subsequence $S_2 = \{a_{22}, a_{23},\ldots\}$ of $S - \{a_{11}\}$ all of whose elements are greater than or equal to a_{11} or all of whose elements are less than or equal to a_{11}. Continuing in this manner, we find an infinite subsequence $S_3 = \{a_{33}, a_{34}, a_{35},\ldots\}$ of $S_2 - \{a_{22}\}$ all of whose elements are greater than or equal to a_{22} or all of whose elements are less than or equal to a_{22}. This process continues, defining a new subsequence S_i at each step. The subsequence $T = \{a_{11}, a_{22}, a_{33},\ldots\}$ has the property that each element a_{ii} is greater than or equal to all elements following it, or less than or equal to all elements following it. Again, by the infinitary pigeonhole principle, there exists a subsequence U of T each of whose elements is greater than or equal to those following it or each of whose elements is less than or equal to those following it. Thus, U is an infinite monotonic subsequence of S.

The set of real numbers **R** is *linearly ordered*. For any two distinct real numbers x and y, either $x < y$ or $y < x$. The forthcoming Dilworth's lemma generalizes the Erdős–Szekeres theorem to partially ordered sets.

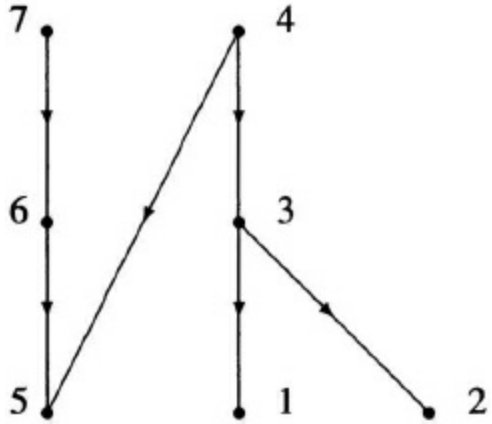
Recall that a partial order on X is a relation on X that is reflexive, antisymmetric, and transitive. We often denote a partial order by $\preceq$ and write $a \preceq b$ when $(a, b) \in \preceq$. Two elements $a, b \in X$ are *comparable* if $a \preceq b$ or $b \preceq a$ and *incomparable* otherwise. For example, the relation $a \preceq b$ if and only if a divides b is a partial order on N.

A partial order $\preceq$ on X is a *total order* (or *linear order*) if every two elements of X are comparable. For example, the usual $\leq$ relation is a linear order on N. If $\preceq$ is a partial order on X, and Y is a subset of X in which every two elements are comparable, then Y is a *chain*. If no two distinct elements of Y are comparable, then Y is an *antichain*. The *length* of a partially ordered set X is the greatest number of elements in a chain of X, and the *width* is the greatest number of elements in an antichain.

EXAMPLE 3.7

Figure 3.3 is the directed graph representation of the partial order

Figure 3.3 A partial order of width 3 and length 3.



$$\preceq = \{(3,1),(4,1),(4,3),(3,2),(4,2),(4,5),(6,5),(7,5),(7,6)\}$$

on the set $X = \{1, 2, 3, 4, 5, 6, 7\}$. The arrows required for reflexivity and transitivity are suppressed in the diagram. The length and width of $\preceq$ are both 3. For instance, $\{1, 3, 4\}$ is a

chain of length 3 and $\{1, 2, 6\}$ is an antichain of size 3.

The length and width of a partial order are related to the size of the underlying set by the following result of R. P. Dilworth.

Dilworth's lemma. In any partial order on a set of $mn + 1$ elements, there exists a chain of length $m + 1$ or an antichain of size $n + 1$.

Proof. Let X be an arbitrary partially ordered set with $mn + 1$ elements. Suppose that X contains no chain of size $m + 1$. Then we may define a function $f : X \to \{1,\ldots,m\}$ with $f(x)$ equal to the greatest number of elements in a chain with greatest element x. By the pigeonhole principle, some $n + 1$ elements of X have the same image under f. By the definition of f, these elements are incomparable; they form an antichain of size $n + 1$.

EXAMPLE 3.8

We consider again the partial order of Figure 3.3. The size of X is $|X| = 7 = 2 \cdot 3 + 1$. Therefore Dilworth's lemma guarantees a chain of $2 + 1$ elements or an antichain of $3 + 1$ elements, and we have remarked that there is a chain of length 3.

Notice the similarity between Dilworth's lemma and the Erdős–Szekeres theorem. In each case, the hypothesis concerns a set of $mn + 1$ elements and the conclusion concerns subsets of sizes $m + 1$ and $n + 1$. These similarities are no coincidence. In fact, the Erdős–Szekeres theorem may be proved as a corollary of Dilworth's lemma. Let $S = \{a_1,\ldots,a_{mn+1}\}$ be a sequence of $mn + 1$ real numbers. Define a partial order $\preceq$ on S by setting $a_i \preceq a_j$ if $a_i \leq a_j$ and $i \leq j$. Dilworth's lemma guarantees a chain of $m+1$ elements (corresponding to a monotonically increasing subsequence of $m + 1$ terms) or an antichain of $n + 1$ elements (corresponding to a monotonically decreasing subsequence of $n + 1$ terms).

Just as the Erdős–Szekeres theorem is a best possible result, so also is Dilworth's lemma. In the exercises, the reader is asked to furnish an example of a partial order on mn elements with length m and width n.

Dilworth's lemma is sometimes easier to apply in the following form.

Dilworth's lemma (alternate form). If a partial order on n elements has length l and width w, then $n \leq lw$.

Proof. Suppose, to the contrary, that $n > lw + 1$. Then, by DilWorth's lemma, there is a chain of length $l + 1$ or an antichain of size $w + 1$; these results contradict the definition of l or w.

As you probably suspect, there is an infinitary version of Dilworth's lemma. The proof is an exercise.

Dilworth's lemma (infinitary version). A partial order on N has an infinite chain or an infinite antichain.

The title Dilworth's lemma suggests that there might be a Dilworth's theorem, which is the case.

Dilworth's theorem. Let X be a partially ordered set with length l and width w. Then X can be partitioned into l antichains or w chains.

Proof We only prove that X can be partitioned into l antichains. For a proof that X can be partitioned into w chains, see [4] or [5].

Define $f : X \to \{1,\ldots,l\}$ by letting $f(x)$ be the maximum number of elements in a chain with greatest element x. The preimage of each $y \in \{1,\ldots,l\}$ is an antichain.

EXAMPLE 3.9

Considering Figure 3.3 again, we find that X may be partitioned into three antichains $\{1,2,5\}$, $\{3,6\}$, $\{4,7\}$, and three chains $\{1,3,4\}$, $\{2\}$, $\{5,6,7\}$.

EXERCISES

3.32 Let $n^2 + 1$ distinct points be given in $\mathbf{R}^2$. Prove that there is a sequence of $n + 1$ points $(x_1, y_1),\ldots,(x_{n+1}, y_{n+1})$ for which $x_1 \le x_2 \le \cdots \le x_{n+1}$ and $y_1 \ge y_2 \ge \cdots \ge y_{n+1}$ or a sequence of $n + 1$ points for which $x_1 \le x_2 \le \cdots \le x_{n+1}$ and $y_1 \le y_2 \le \cdots \le y_{n+1}$.

3.33 Give an example of a partial order on mn elements with length m and width n.

3.34 (Putnam Competition, 1967) Let $0 < a_1 < a_2 < \cdots < a_{mn+1}$ be $mn + 1$ integers. Prove that you can select either $m + 1$ of them no one of which divides any other, or $n + 1$ of them each dividing the following one.

Hint: Apply Dilworth's lemma.

3.35 For any n^2+1 closed intervals of $\mathbf{R}$, prove that $n + 1$ of the intervals share a point or $n + 1$ of the intervals are disjoint.

Hint: Let $\alpha \preceq \beta$ if the closed interval α is entirely to the left of the closed interval β. Apply Dilworth's lemma.

3.36 Prove the infinitary Dilworth's lemma.

3.37 Prove the infinitary Erdős–Szekeres theorem as a corollary of the infinitary Dilworth's lemma.

3.38 Let $\preceq_1$ and $\preceq_2$ be two total orders on a set of size $n^2 + 1$. Show that there is a subset of size $n + 1$ on which $\preceq_1$ and $\preceq_2$ totally agree or totally disagree.

3.39 Prove that if $2n - 1$ total orders are given on $m^{2^{2n-2}} + 1$ points, then some $m + 1$ points are totally ordered by n agreeing orders.

3.6 Subsets

Let $X(t)$ be the collection of subsets of the t-element set N_t, and let $\subseteq$ be the containment partial order on $X(t)$. For instance, if $t = 3$, then $X(t)$ consists of eight elements: θ, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, and $\{1,2,3\}$. Some examples of containment are $\{1, 2\} \subseteq \{1, 2, 3\}$, $\theta \subseteq \{3\}$, and $\{1, 3\} \subseteq \{1,$

3}. The size of $X(t)$ is 2^t. What are the length and width of $\subseteq$? The length is $t + 1$, because the longest chains start with θ and include one new element at each step until all t elements are included. The width of $\subseteq$ is the subject of Sperner's theorem. Remember that an antichain of $X(t)$ is a collection of subsets of N_t none of which contains another.

Emanuel Sperner (1905–1980) proved the following result in 1928. We give a simpler proof essentially due to D. Lubell. See [9]. For a proof of Sperner's theorem using the probabilistic method, see [2].

Sperner's theorem. An antichain of subsets of N_t (under the usual $\subseteq$ order) has at most $\binom{t}{\lfloor t/2 \rfloor}$ elements. Furthermore, the $\binom{t}{\lfloor t/2 \rfloor}$ subsets of size $[t/2]$ form an antichain.

Note that Sperner's theorem tells us that the width of the partially ordered set $X(t)$ is $\binom{t}{\lfloor t/2 \rfloor}$.

Proof Let $A = \{A_1,\ldots,A_m\}$ be an antichain of subsets of N_t with $|A_i| = \alpha_i$ at for $1 \le i \le m$. For each i, the set A_i is contained in exactly $\alpha_i!(t - \alpha_i)!$ chains of length $t + 1$. (Such chains commence with the empty set, add one element at a time until A_i is exhausted, then add one element at a time until the complement of A_i is exhausted.) Because these chains are distinct and there are $t!$ chains of length $t + 1$ altogether,

$$\sum_{i=1}^{m} \alpha_i!(t - \alpha_i)! \le t!.$$

Dividing this inequality by $t!$ we obtain

$$\sum_{i=1}^{m} \binom{t}{\alpha_i}^{-1} \le 1.$$

Since $\binom{t}{k}$ is maximized when $k = [t/2]$, it follows that

$$m\binom{t}{\lfloor t/2 \rfloor}^{-1} \le \sum_{i=1}^{m} \binom{t}{\alpha_i}^{-1} \le 1.$$

Therefore

$$m \le \binom{t}{\lfloor t/2 \rfloor}.$$

What are the antichains of $X(t)$ with $\binom{t}{\lfloor t/2 \rfloor}$ elements? Equality in the above relation can be attained only if $\binom{t}{\lfloor t/2 \rfloor}^{-1} = \binom{t}{\alpha_i}^{-1}$ for each α_i. If t is even, this forces $\alpha_i = t/2$. If t is odd, then α_i can equal $(t - 1)/2$ or $(t + 1)/2$. We now prove that if t is odd, then all elements of the antichain are size $(t - 1)/2$ or all are size $(t + 1)/2$. The proof is essentially due to László Lovász.

Theorem. Let A be an antichain of $X(t)$ containing $\binom{t}{\lfloor t/2 \rfloor}$ elements. If t is even, then A is the collection of all subsets of N_t of size $t/2$. If t is odd, then A is the collection of all subsets of size $(t - 1)/2$ or the collection of all subsets of size $(t + 1)/2$.

Proof. We have already demonstrated the t even case. Suppose that $t = 2u + 1$. As each maximal chain

in $X(t)$ contains exactly one element of A, if U is a subset of size u, V is a subset of size $u + 1$, and $U \subseteq V$, then A contains exactly one of U and V. Suppose that U is a subset of size u contained in A, and U' is any other subset of size u. Then there is a sequence of subsets

$$U = U_1, V_1, U_2, V_2, \ldots, V_{n-1}, U_n = U'$$

beginning with U and ending with U', whose sizes alternate between u and $u + 1$, and such that V_i contains U_i, and U_{i+1} for each i. It follows that U' is an element of A. Because U' was arbitrary, A contains all subsets of size u. A similar argument shows that if A contains at least one subset of size $u + 1$, then A contains every subset of size $u + 1$.

We are now ready to look at relations which are not transitive. In Chapter 4 we begin by discussing graphs, where the relations are merely reflexive and symmetric. The theorems are more difficult to prove in this more general setting—and the analysis of best possible results is *much* more difficult.

EXERCISES

3.40 Let $a_1,\ldots,a_n, b \in \mathbf{R}$, with all $a_i \geq 1$. Show that the maximum number of sums $\pm a_1 \pm a_2 \pm \cdots \pm a_n$ in the open interval $(b, b + 2)$ is $\binom{n}{\lfloor n/2 \rfloor}$.

3.41 Let $a_1,\ldots,a_n$ be positive real numbers. Show that the maximum number of equal sums $\in_1 a_1 + \cdots + \in_n a_n$ ($\in_i = 0$ or 1) is $\binom{n}{\lfloor n/2 \rfloor}$.

See [3] and [26] for a discussion of the Littlewood–Offord problem concerning the number of sums $\Sigma^n_{i=1} \in_i z_i$ ($\in_i = \pm 1$ and $|z_i| \geq 1$) lying inside any given circle in the complex plane.

Notes

Johann Peter Gustav Lejeune Dirichlet (1805–1859) was the first mathematician to explicitly use the pigeonhole principle in proofs. He referred to it as the "drawer principle."

The word "graph" was first used in mathematics in an 1877 paper by James Sylvester (1814–1897). In 1936 Dénes König (1884–1944)> wrote the first book on graph theory, *Theorie der endlichen and unendlichen Graphen.*

The special case of Turán's theorem (1941) was proved by W. Mantel in 1907.

The Erdős–Szekeres theorem was proved in 1935 and may be regarded as a sort of proto-Ramsey theorem (even though Ramsey's theorem was proved in 1930).

According to the Erdős–Szekeres theorem, every permutation of the numbers 1,…,10 contains a monotonic subsequence of length four. But this result does not hold for permutations of the numbers 1, …,9; there are many permutations of 1,…,9 that do not have a monotonic subsequence of length four. The question of exactly how many is answered by the theory of Young tableaux. For a discussion of Young tableaux, see [19] and [26]. We give a few details here.

Let $n = \lambda_1 + \cdots + \lambda_n$. A *Young tableau of shape* $\lambda_1 + \cdots + \lambda_n$ is a Ferrers diagram of shape $(\lambda_1, \ldots,\lambda_n)$ in which each dot has been replaced by a different integer from the set $\{1,\ldots,n\}$. The number n is the *order* of the tableau. The positions in a tableau are called *cells*. A *standard* tableau is one in which the integers increase in every column and in every row.

Figure 3.4 shows an example of a standard Young tableau with $n = 9 = 4+3+2$.

Figure 3.4 A standard Young tableau of shape 4 + 3 + 2.

1	3	4	8
2	5	9	
6	7		

How many standard Young tableaux have this shape? The answer is given by the *hook-length formula*. We define the *hook-length* of a cell to be one more than the number of cells to its right and below it. Figure 3.5 shows the hook-lengths of the cells of the tableau of Figure 3.4.

Figure 3.5 The hook-lengths of the cells of a tableau.

6	5	3	1
4	3	1	
2	1		

The hook-length formula says that the number of standard Young tableaux of a given shape is equal to $n!$ divided by the product of the hook-lengths. Thus, the number of standard Young tableaux of shape 4 + 3 + 2 is

$$\frac{9!}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 3 \cdot 2 \cdot 1 \cdot 1 \cdot 1} = 168.$$

There are 30 partitions of the number 9. For each such partition λ, let f_λ be the number of standard tableaux of shape λ. If you compute these numbers (via the hook-length formula), you may be surprised to find that

$$\sum_\lambda f_\lambda^2 = 9!.$$

This identity (for any positive integer n) is known as *Schur's formula*. It is used in the theory of representations of the symmetric group.

Schur's formula. For each partition λ of n, let f_λ be the number of standard Young tableaux of shape λ. Then

$$\sum_\lambda f_\lambda^2 = n!.$$

Schur's formula indicates that there is a correspondence between pairs of standard Young tableaux of identical shape and permutations of n. This correspondence is effected by the *Robinson–Schensted algorithm*.

We give an example of the algorithm. Let $n = 9$ and

$$\sigma = 583276491.$$

We will construct an ordered pair (P, Q) of standard Young tableaux (of the same shape) corresponding to σ.

The first task is to construct P. (We will construct Q later.) We read the permutation σ from left to right and construct P step by step. The 5 is placed in the top left position of the tableau, and the 8, being greater than 5, is placed below:

5

8

Now we come to the 3. Being less than 5, the 3 "bumps" the 5 to the right and takes its place:

3 5

8

Likewise, 2 is less than 3, so it bumps the 3 to the right (arid the 5 along with it) and takes its place:

2 3 5

8

Now comes the 7. Because 7 is less than 8, it is inserted into the second row, bumping the 8 to the right. Then the 6 bumps the 7 (and the 8 along with it):

2 3 5

6 7 8

Next, the 4 bumps the entire second row to the right, and the 8 is bumped up to the first row:

2 3 5 8

4 6 7

Finally, the 9, being greater than 4, is placed in the third row, and the 1 bumps the first row to the right:

1 2 3 5 8

4 6 7

9

This completes the tableau P. The tableau Q consists of the numbers 1,…,9 in a tableau of the same shape as P and in the order in which new positions were occupied in P. Figure 3.6 shows P and Q.

Figure 3.6 The tableaux P and Q corresponding to $\sigma = 583276491$.

1	2	3	5	8
4	6	7		
9				

1	3	4	7	9
2	5	6		
8				

In the Robinson–Schensted correspondence, the number of columns of the tableaux P and Q is equal to the length of a longest decreasing subsequence of the permutation, and the number of rows is equal to the length of a longest increasing subsequence. In our example, the tableaux have five columns and three rows, and indeed, a longest decreasing subsequence of σ has five terms and a longest increasing subsequence has three terms.

To calculate the number of permutations of $\{1,\ldots,9\}$ with no monotonic subsequence of length four, we use the hook-length formula to find the number of standard Young tableaux of shape 3 + 3 + 3:

$$\frac{9!}{5 \cdot 4 \cdot 4 \cdot 3 \cdot 3 \cdot 3 \cdot 2 \cdot 2 \cdot 1} = 42.$$

Since the desired permutations correspond to ordered pairs of standard Young tableaux of this shape, the number of such permutations is $42^2 = 1764$.

We leave five problems for you to ponder:

1. How does the reverse direction of the Robinson–Schensted algorithm work?

2. In the Robinson–Schensted correspondence, show that the number of columns of the tableaux P and Q is equal to the length of the longest decreasing subsequence of σ. Show that the number of

rows is equal to the length of the longest increasing subsequence of σ.

3. If the permutation σ corresponds to the pair of tableaux (P, Q), show that σ^{-1} corresponds to the pair (P, Q).

4. A permutation $\sigma \in S_n$ is called an *involution* if $\sigma = \sigma^{-1}$. Show that the number of involutions in S_n, equals the number of standard tableaux of order n. Recall that these were counted in Exercise 2.69.

5. Show that the number of standard Young tableaux of shape $2 \times n$ is given by the Catalan number C_n.

CHAPTER 4

RAMSEY THEORY

The Erdös–Szekeres theorem and Dilworth's lemma guarantee the existence of particular substructures of certain combinatorial configurations. Large disordered structures contain ordered substructures. We continue this theme by presenting two cornerstones of Ramsey theory: Ramsey's theorem on graphs and van der Waerden's theorem on arithmetic progressions. In the process we discuss related results, including Schur's theorem on equations. We also investigate bounds and asymptotics of Ramsey numbers using techniques from number theory and probability. The central pursuit is always to find ordered substructures of large disordered structures. We want to find order in randomness.

4.1 Ramsey's theorem

The following problem appeared in the 1953 William Lowell Putnam Mathematical Competition:

> Six points are in general position in space (no three in a line, no four in a plane). The fifteen line segments joining them in pairs are drawn and then painted, some segments red, some blue. Prove that some triangle has all its sides the same color.

The description of the six points in general position and the segments joining them in pairs is just another way of defining the graph K_6. We introduce a few more graph theory terms. A *coloring* of the set of edges of a graph G is a function $f : E(G) \to S$, where S is a set of *colors*. A coloring partitions $E(G)$ into *color classes*. If f is constant, then G is *monochromatic*.

Now we may rephrase the Putnam question as follows: If each edge of K_6 is colored either red or blue, then there is a monochromatic subgraph K_3 (a triangle). We note that the coloring may be done in an arbitrary manner. In fact, because K_6 has $\binom{6}{2} = 15$ edges, there are 2^{15} possible red–blue colorings of the edges of K_6. The claim is that every one of these 32,768 colorings yields a monochromatic K_3. (We assume that the vertices of K_6 are labeled, so we can distinguish between differently labeled isomorphic graphs, and that all 15 edges can be the same color, a possibility disallowed in the Putnam problem as stated.)

Here is a simple solution to the problem using the pigeonhole principle. Choose any vertex v of K_6. By the pigeonhole principle, some three of the five edges emanating from v are the same color. Without loss of generality, suppose v is joined by red edges to vertices x, y, z. If any of the edges xy, yz, or xz is red, then there is a red triangle (vxy, vyz, or vxz). However, if each of these edges is blue, then xyz is a blue triangle.

A special notation has been introduced to state results of this type. We write

(4.1) $K_6 \longrightarrow (K_3)_2$

to indicate that every 2-coloring of the edges of K_6 yields a monochromatic K_3. This relation may also be written

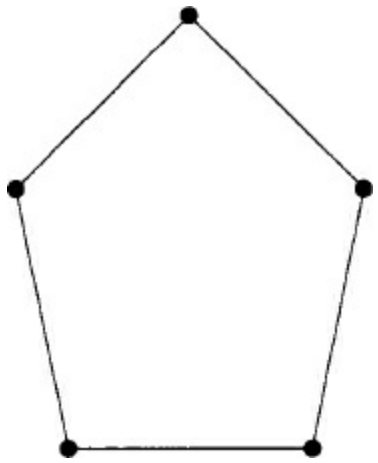
(4.2) $K_6 \underset{2}{\longrightarrow} K_3.$

Similarly, we write

(4.3) $K_5 \not\longrightarrow (K_3)_2$

to say that there is a 2-coloring of K_5 with no monochromatic K_3. It is equivalent to say that there is a graph G on five vertices such that neither G nor $\overline{G}$ contains a K_3. Such a graph is exhibited in Figure 4.1.

Figure 4.1 A graph G such that neither G nor $\overline{G}$ contains K_3.



In general, we write

(4.4) $K_n \longrightarrow (K_m)_2$

to indicate that every 2-coloring of the edges of K_n, yields a monochromatic K_m. In 1930 F. Ramsey established the existence of such a K_n, for each m. Unlike the authors of the Putnam problem, we prefer green–red colorings to red–blue colorings.

Ramsey's theorem (1930). Given $a, b \geq 2$, there exists a least integer $R(a,b)$ with the property that every green–red coloring of the edges of the complete graph on $R(a,b)$ vertices yields a green K_a or a red K_b. Furthermore,

(4.5) $R(a,b) \leq R(a-1,b) + R(a,b-1),$

for all $a, b \geq 3$.

Proof. We employ induction on a and b. The basis of the induction consists of the statements $R(a, 2) = a$ and $R(2, b) = b$. These are trivial. In the first assertion, if we 2-color K_a and any edge is red, then we obtain a red K_2, while if no edge is red, then we obtain a green K_a. Thus $R(a, 2) \leq a$. Equality follows from the fact that an all-green-colored K_{a-1} contains neither a green K_a nor a red K_2. The second assertion is proved similarly. Now, assuming the existence of $R(a - 1, b)$ and $R(a, b - 1)$, we will show that $R(a, b)$ exists. Let G be the complete graph on $R(a - 1, b) + R(a, b - 1)$ vertices, and let v be an arbitrary vertex of G. By the pigeonhole principle, at least $R(a - 1, b)$ green edges or at

least $R(a,b–1)$ red edges emanate from v. Without loss of generality, suppose that v is joined by green edges to a complete subgraph on $R(a-1, b)$ vertices. By definition of $R(a-1, b)$, this subgraph must contain a red K_b or a green K_{a-1}. In the latter case, the green K_{a-1} and v, and all the edges between the two, constitute a green K_a. We have shown that G contains a green K_a or a red K_b. Therefore, $R(a, b)$ exists and satisfies

$$R(a,b) \leq R(a-1,b) + R(a,b-1).$$

The values of $R(a, b)$ are called *Ramsey numbers*. Very few nontrivial Ramsey numbers (with a or b greater than 2) have been determined. The fact that we have proved that the Ramsey numbers exist but we do not know their values illustrates one disadvantage of existential proofs.

By definition, $R(m, m)$ is the least positive integer n for which $K_n \rightarrow (K_m)_2$. The values $R(m, m)$ are called *diagonal Ramsey numbers* because they appear on the main diagonal of a table of Ramsey numbers. We know one diagonal Ramsey number already: $R(3, 3) = 6$.

We note that $R(a, b) = R(b, a)$ for all $a, b \geq 2$, as the roles of the two variables a and b are symmetric. Furthermore, we have noted that $R(a, 2) = a$ for all a. We have already proved that $R(3, 3) = 6$, but a second proof is furnished by the two observations just made and the inequality $R(a, b) \leq R(a-1, b) + R(a, b-1)$. Thus, $R(3, 3) \leq R(3, 2) + R(2, 3) = 3 + 3 = 6$. The lower bound $R(3, 3) > 5$ is verified by construction as before.

EXAMPLE 4.1 Confusion graph

The *confusion graph* is defined as follows. Suppose that the vertices of the 5-cycle C_5 are a, b, c, d, e (in cyclic order) and these vertices represent symbols transmitted over a noisy channel. Adjacent symbols are said to be *confusable;* each is easily mistaken for the other. Nonadjacent symbols are not confusable. Thus, c and d are confusable while c and e are not. The independence number $\alpha(G)$ is the maximum number of nonconfusable symbols in $V(G)$. It is easy to see that $\alpha(C_5) = 2$, as illustrated by the pair a, d. For two finite graphs G and H, we define a new graph called the *strong product* $G \boxtimes H$ as the set $V(G) \times V(H)$, with (g, h) adjacent to (g', h') if and only if g is adjacent to or equal to g' and h is adjacent to or equal to h'. We think of $\alpha(G \boxtimes H)$ as the maximum number of nonconfusable ordered pairs in the set $V(G) \times V(H)$, where nonconfusability means nonconfusability in at least one coordinate. (We are assuming that a symbol is confusable with itself.)

If A is a nonconfusable subset of $V(G)$ and B is a nonconfusable subset of $V(H)$, then $A \times B$ is a nonconfusable subset of $V(G) \times V(H)$. Therefore $\alpha(G)\alpha(H) \leq \alpha(G \boxtimes H)$. Ramsey's theorem furnishes a strict upper bound: $\alpha(G \boxtimes H) < R(\alpha(G) + 1, \alpha(H) + 1)$. For suppose the upper bound is attained by a subset $A \times B$ of $V(G) \times V(H)$. Color an edge green if there is nonconfusability in the first coordinate and red if there is nonconfusability in the second coordinate. (If nonconfusability holds in both coordinates, then we color the edge green.) Ramsey's theorem guarantees that A has at least $\alpha(G) + 1$ points or that B has at least $\alpha(H) + 1$ points—both contradictions. Putting our lower and upper bounds together we obtain $4 = \alpha(C_5)^2 \leq \alpha(C_5 \boxtimes C_5) < R(\alpha(C_5) + 1, \alpha(C_5) + 1) = R(3, 3) = 6$. Is the value of $\alpha(C_5 \boxtimes C_5)$ 4 or 5?

EXERCISES

4.1 For the confusion graph C_5, show that $\alpha(C_5 \boxtimes C_5) = 5$.

4.2 A *tournament* is a complete directed graph. Use Ramsey's theorem to show that for every n there exists an $f(n)$ such that every tournament on $f(n)$ vertices contains a transitive subtournament on n vertices.

4.3 Prove that every 2-coloring of the edges of K_6 yields two monochromatic triangles.

Hint: Assign to each pair of edges incident at a vertex a score of +2 if they are the same color and –1 if not.

4.2 Generalizations of Ramsey's theorem

What we can do with two colors, we can do with an arbitrary number, as the following generalization of Ramsey's theorem shows. All the theorems of this section were proved by Frank Ramsey in the original 1930 paper. See Notes.

Ramsey's theorem for multiple colors. For any $c \geq 2$ and $a_1,\ldots,a_c \geq 2$, there exists a least integer $R(a_1,\ldots,a_c)$ with the following property: If the edges of the complete graph on $R(a_1,\ldots,a_c)$ vertices are partitioned into color classes $A_1,\ldots,A_c$, then for some i there exists a complete graph on a_i, vertices all of whose edges are color A_i.

Proof. The case $c = 2$ is covered by our previous version of Ramsey's theorem. Suppose $R(a_1,\ldots,a_{c-1})$ exists for all $a_1,\ldots,a_{c-1} \geq 2$. We claim $R(a_1,\ldots,a_c)$ exists and satisfies $R(a_1,\ldots,a_c) \leq R(R(a_1,\ldots,a_{c-1}), a_c)$. A c-coloring of the complete graph on $R(R(a_1,\ldots,a_{c-1}, a_c)$ vertices may be regarded as a 2-coloring with colors $\{A_1,\ldots,A_{c-1}\}$ and A_c. Such a coloring contains a complete graph on a_c vertices colored A_c or a $(c-1)$-colored complete graph on $R(a_1,\ldots,a_{c-1})$ vertices, in which case the induction hypothesis holds. In either case, we obtain a complete subgraph on the required number of vertices.?

EXAMPLE 4.2

We use the c-color Ramsey theorem to prove a weak version of Dilworth's lemma. Recall that Dilworth's lemma states that every partial order on $mn + 1$ elements contains a chain of length $m + 1$ or an antichain of size $n + 1$. For k sufficiently large, we define a coloring of the complete graph on the vertex set $X - \{x_1,\ldots,x_k\}$ as follows: Assuming $i < j$, color edge x_ix_j blue if x_i and x_j are incomparable; green if $x_i \leq x_j$; and red if $x_i \geq x_j$. (Some edges may be colored in two ways, but it won't matter.) Now if $k = R(n + 1, m + 1, m + 1)$, then we are guaranteed a blue K_{n+1} (corresponding to an antichain of size $n + 1$), a green K_{m+1} (corresponding to a chain of x_i with increasing subscripts), or a red K_{m+1} (corresponding to a chain of x_i, with decreasing subscripts). Thus we have a weak version of Dilworth's lemma. It is true that the best possible value $mn + 1$ has been replaced by the presumably much larger value $R(n + 1, m + 1, m + 1)$. However, we have gained information about the increasing or decreasing nature of the subscripts of the x_i. It would be unreasonable to expect that the best possible value $mn + 1$ would be obtained by this

proof, because Dilworth's lemma assumes a transitive relation while Ramsey's theorem does not. Thus Ramsey's theorem is more general than Dilworth's lemma.

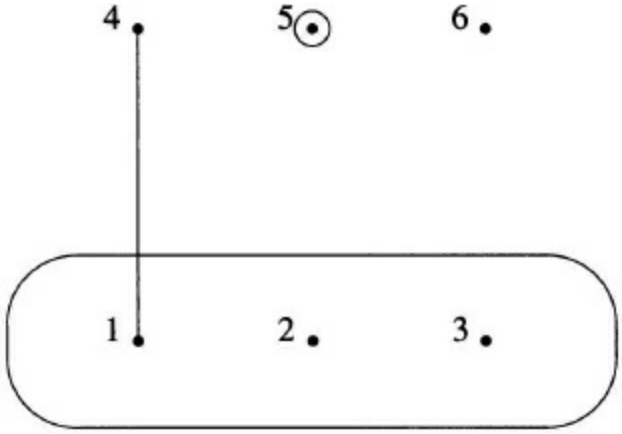
We write

(4.6) $K_n \to (K_m)_c$

to indicate that every c-coloring of K_n, yields a monochromatic K_m. The c-color Ramsey numbers $R(a_1,\ldots,a_c)$ satisfy certain trivial relations, e.g., they are symmetric in the c variables. Furthermore, $R(a_1,\ldots,a_{c-1}, 2) = R(a_1,\ldots,a_{c-1})$ for all a_i, because either there is an edge colored A_c, or else all edges are colored from the set $\{A_1,\ldots,A_{c-1}\}$.

A *hypergraph* H of *order* n is a collection of nonempty subsets of an n-set S of *vertices*. The elements of H are called *edges*, corresponding to the graphical case in which edges are two-element subsets. A hypergraph is *t-uniform* if all edges have cardinality t. The *complete* t-uniform hypergraph of order n is the collection $[S]^t$ of all t-subsets of S. One may visualize and draw hypergraphs with edges represented by ovals (cardinality > 2), lines (cardinality = 2), and circles (cardinality = **1**), as in Figure 4.2.

Figure 4.2 A hypergraph with edges {1,2,3}, {1,4}, and {5}.



It is now possible to state the most general version of Ramsey's theorem in its natural hypergraph setting. We sometimes write $[N_n]^t$ as $[n]^t$.

Ramsey's theorem for hypergraphs. Let $c \geq 2$ and $a_1,\ldots,a_c \geq t \geq 2$. There exists a least integer $R(a_1,\ldots,a_c; t)$ with the following property: Every c-coloring of the complete t-uniform hypergraph $[R(a_1,\ldots,a_c;t)]^t$ with colors $A_1,\ldots,A_c$ yields a complete t-uniform hypergraph on a_i vertices in color A_i, for some i.

Proof The generalization from 2-colorings to c-colorings works as it did in the generalization from the two-color to the c-color Ramsey theorem. We leave the argument as an exercise. We know that $R(a_1, a_2; 2)$ exists for all $a_1, a_2 \geq 2$. Let us assume that $R(a_1, a_2; t-1)$ exists for all $a_1, a_2 \geq 2$ and that $R(a_1 - 1, a_2; t)$ and $R(a_1, a_2 - 1; t)$ exist. We claim that $R(a_1, a_2; t)$ exists and satisfies $R(a_1, a_2; t) \leq n$, where

$$n = 1 + R(R(a_1 - 1, a_2; t), R(a_1, a_2 - 1; t); t - 1).$$

Suppose $[N_n]^t$ has been green–red colored, and let v be one of its vertices. We generate an *induced* 2-

coloring of $[N_n - \{v\}]^{t-1}$ by assigning to each $(t-1)$-set A of $N_n - \{v\}$ the color which has been assigned to the t-set $A \cup \{v\}$. By definition of n, we know that $[N_n - \{v\}]^{t-1}$ contains a green $[R(a_1 - 1, a_2; t)]^{t-1}$ or a red $[R(a_1, a_2 - 1; t)]^{t-1}$. Without loss of generality, suppose there is a green $[R(a_1 - 1, a_2; t)]^{t-1}$ on vertex set A. By definition, $[A]^t$ contains a red $[a_2]^t$ or a green $[a_1 - 1]^t$. In the latter case, $[A \cup \{v\}]^t$ contains a green $[a_1]^t$. We have shown that $[N_n]^t$ contains a green $[a_1]^t$ or a red $[a_2]^t$, as required.

An application of the theorem to convex sets occurs in the exercises. Now we discuss infinitary versions of the Ramsey theorems. We write

(4.7) $$K_\infty \xrightarrow[c]{} K_\infty$$

to indicate that every c-coloring of the complete countably infinite graph yields a monochromatic complete countably infinite subgraph. Similarly, in the hypergraph setting, the *infinite t-uniform complete graph* $K_\infty^{(t)}$ consists of a countably infinite set and all possible t-element subsets. We write

(4.8) $$K_\infty^{(t)} \xrightarrow[c]{} K_\infty^{(t)}$$

to indicate that every c-coloring of the t-uniform complete infinite graph yields a monochromatic t-uniform complete infinite subgraph.

Ramsey's theorem for infinite graphs. For every $c \geq 2$, we have $K_\infty \xrightarrow[c]{} K_\infty$.

Proof. Define $f : N \to \{1,\ldots,c\}$ as follows. Let $n = 1$ and $X_n = V(K_\infty)$. Choose $x_n \in X_n$, and let $A_i = \{v \in X_n : \text{edge } vx_n \text{ is color } i\}$. By the infinitary pigeonhole principle, some A_i is infinite. Let $X_{n+1} = A_i$, and define $f(n) = i$ accordingly. Replace n by $n + 1$ and repeat this process.

This recursive procedure defines the function f. Some $f^{-1}(j)$ is infinite and the complete graph on vertex set $\{x_n : n \in f^{-1}(j)\}$ is monochromatic.

Ramsey's theorem for infinite hypergraphs. For every $t \geq 2$, $c \geq 2$, we have $K_\infty^{(t)} \xrightarrow[c]{} K_\infty^{(t)}$.

The infinite hypergraph Ramsey theorem includes the infinite graph Ramsey theorem as a special case (when $t = 2$). The proof of the infinite hypergraph Ramsey theorem is left as an exercise.

We close this section by indicating how Ramsey's theorem for infinite graphs implies Ramsey's theorem for finite graphs. The technique for doing this, the "compactness principle," is used throughout combinatorics. Assuming the truth of the infinite graph Ramsey theorem, we prove the finite graph Ramsey theorem by contradiction. Suppose that there exists a k for which $R(k, k)$ does not exist. For each $i \geq k$, let f_i be a 2-coloring of K_i without a monochromatic K_k. We assume the K_i, are nested: $K_k \subseteq K_{k+1} \subseteq K_{k+2} \subseteq \cdots$. By the infinitary pigeonhole principle, there exists an infinite subset of functions $\{f_i^k\} \subseteq (f_i\}$ which agree on K_k. Similarly, there is an infinite subset of functions $\{f_i^{k+1}\} \subseteq \{f_i^k\}$ agreeing on K_{k+i}, etc. This process yields an infinite 2-coloring of K_∞ without a monochromatic

K_k, contradicting the infinite graph Ramsey theorem. Therefore, $R(k, k)$ exists for each k, which implies that $R(a_1, a_2)$ exists, since it must satisfy the inequality $R(a_1, a_2) \leq R(\max\{a_1, a_2\}, \max\{a_1, a_2\})$. In the same manner, the finite hypergraph Ramsey theorem for an arbitrary number of colors is proven from the infinite hypergraph Ramsey theorem for an arbitrary number of colors.

EXERCISES

4.4 Prove the following result of Erdős and Szekeres (1935): For every m, there exists a least integer $n(m)$ such that any set of $n(m)$ points in the plane contains m points which determine a convex m-gon.

Hint: $n(m)$ satisfies $n(m) \leq R(5, m; 4)$. Actually, Erdős and Szekeres proved that $n(m) \geq 2^{m-2} + 1$ and conjectured that $n(m) = 2^{m-2}$ 1. The determination of $n(m)$ remains an open problem.

4.5 Prove that among infinitely many points in the plane there are infinitely many collinear points or infinitely many points no three of which are collinear. Prove also the three-dimensional analog of this problem: Among infinitely many points in R^3 there is an infinite planar subset or an infinite subset containing no four coplanar points.

4.6 A c-coloring of the edges of a graph is *surjective* if all c colors are used. For $a \geq b \geq 1$, let $P(a, b)$ be the following proposition: Every surjective a-coloring of the countably infinite complete graph yields a surjectively b-colored infinite complete subgraph.

(a) Show that $P(a, b)$ is true if $b = 1$, $b = 2$, or $a = b$. It is conjectured that these are the only a and b for which $P(a, b)$ is true.

(b) Show that $P(10, 8)$ and $P(46, 15)$ are false.

4.7 Prove that if $K_{\infty,\infty}$ is 2-colored, there exists a monochromatic $K_{\infty,\infty}$. Interpret this result as a proposition about 2-colorings of the infinite square lattice.

4.8 (Putnam Competition, 1988) (a) If every point of the plane is painted one of three colors, do there necessarily exist two points of the same color exactly one inch apart?

(b) What if "three" is replaced by "nine"?

Justify your answers.

The answer to (a) is yes and the answer to (b) is no. The minimum number of colors necessary to force the conclusion in part (a) is not known. See [18].

4.9 (Putnam Competition, 1994) Show that if the points of an isosceles right triangle of side length 1 are each colored with one of four colors, then there must be two points of the same color which are at distance at least $2 - \sqrt{2}$ apart.

4.10 Find an infinite graph G with the following three properties:

(1) G contains no K_4.

(2) The addition of any edge to G completes a K_4.

(3) There is a 2-coloring of the edges of G with no monochromatic K_3.

In 1988 Joel Spencer used the probabilistic method to prove the existence of a finite graph G with properties (1) and (2) *without* property (3) and with fewer than $3 \cdot 10^9$ vertices. His result answers a question of Erdős, who asked whether there exists such a graph with at most

10^{10} vertices. See J. Spencer, Three hundred million points suffice, *Journal of Combinatorial Theory (A)* **49** (1988), 210–217, and Erratum, *Journal of Combinatorial Theory (A)* **50** (1989), 323.

For $m, n \geq 3$ and $p > \max\{m, n\}$, the *Folkman number* $F(m, n; p)$ is defined as the minimum number of vertices in a graph G with the properties (1) the largest complete graph contained in G has p vertices and (2) any green–red coloring of the edges of G yields a green K_m or a red K_n. In 1967 Jon Folkman proved the existence of these numbers. Spencer's construction proves $F(3, 3; 3) < 3 \cdot 10^9$. Other than the Ramsey numbers (i.e., when $p = R(m, n)$), the only known Folkman numbers are $F(3, 3; 5) = 8$ and $F(3, 3; 4) = 15$. See [10, p. 1373].

4.3 Ramsey numbers, bounds, and asymptotics

Until now our results have been purely existential. We have shown that sufficiently large structures contain desired nonrandom substructures. But how large is sufficiently large? In general, this quantification question is extremely difficult, and unsolved problems abound. We present a few calculations and proofs in this section and summarize the scant amount of information known about Ramsey numbers.

We have already shown that $R(3, 3) = 6$. Let us try to evaluate the next more complicated Ramsey number, $R(3, 4)$. To obtain an upper bound, we use the inequality $R(a, b) \leq R(a - 1, b) + R(a, b - 1)$. Thus $R(3, 4) \leq R(3, 3) + R(2, 4) = G + 4 = 10$. However, $R(3, 4)$ turns out to be 9, not 10. For suppose there is a green–red coloring of K_9 which has no green K_3 and no red K_4. Because $R(2, 4) = 4$ and $R(3, 3) = 6$, each vertex must be incident with exactly three green edges and five red edges. But this means that the sum of the degrees of the vertices of the green subgraph is $9 \cdot 3 = 27$, contradicting the fact that the sum of degrees is always even (the Handshake Theorem). Hence $R(3, 4) \leq 9$. In the exercises, the reader is asked to furnish a 2-coloring of K_8 containing no green K_3 and no red K_4, thereby proving $R(3, 4) = 9$.

The Ramsey number $R(3, 5)$ is evaluated easily: $R(3, 5) \leq R(3, 4) + R(2, 5) = 9 + 5 = 14$. In the exercises, the reader is asked to find a 2-coloring of K_{13} that shows $R(3, 5) > 13$, thus establishing $R(3, 5) = 14$.

Next we determine $R(4, 4)$. We have the upper bound $R(4, 4) \leq R(4, 3) + R(3, 4) = 9 + 9 = 18$, and 18 turns out to be the value of $R(4, 4)$. To prove this, we need a green-red coloring of K_{17} containing no monochromatic K_4. In general, colorings which establish lower bounds tend to look locally random. However, they must contain quite a bit of structure so that they can be manipulated and analyzed. Such *pseudorandom* constructions are employed throughout combinatorics.

Let us assume that the vertices of K_{17} are labeled with the residue classes modulo 17: 0, 1, 2,…,16. An edge ij is colored green or red according to the *quadratic character* of $i–j$ modulo 17. The 16 nonzero residues fall into two classes, quadratic residues and quadratic nonresidues. The set of quadratic residues modulo 17 is

$$R_{17} = \{x^2 : x \in \mathbf{Z}_{17}^*\} = \{1, 4, 9, 16, 8, 2, 15, 13\},$$

the set of quadratic nonresidues is

$N_{17} = \{3, 5, 6, 7, 10, 11, 12, 14\}$.

Note that R_{17} is the range of the homomorphism $f : \mathbf{Z}^*_{17} \to \mathbf{Z}^*_{17}, x \mapsto x^2$. Both R_{17} and N_{17} are closed under multiplication by –1 (because $-1 = 16 \in R_{17}$), so $i - j$ has the same quadratic character as $j - i$. Edge ij is colored green if $i - j \in R_{17}$ and red if $i - j \in N_{17}$. Suppose that there is a monochromatic K_4 on vertices a, b, c, d. Note that the coloring is translation invariant: $(i + k) - (j + k) = i - j$. Therefore we may assume that $a = 0$. Multiply each vertex by b^{-1} (the multiplicative inverse of b), and note that either no edge changes color (if $b \in R_{17}$) or else every edge changes color (if $b \in N_{17}$). The reason for this is that $b^{-1}i - b^{-1}j = b^{-1}(i - j)$. In either case, we now have a monochromatic K_4 on vertices 0, 1, cb^{-1}, db^{-1}. Now, because 1 is a quadratic residue, the other differences $cb^{-1}\ db^{-1}$, $cb^{-1} - 1$, $db^{-1} - 1$, $db^{-1} - cb^{-1}$ must be quadratic residues. Upon inspection of the elements of R_{17}, we see that this is impossible. Therefore $R(4, 4) > 17$, and we conclude that $R(4, 4) = 18$.

The other two-color Ramsey numbers $R(a, b)$ are considerably more difficult to evaluate. The above construction involving quadratic residues was discovered in 1955 by R. E. Greenwood and A. M. Gleason. Although it gives the exact Ramsey number in the case of $R(4, 4)$, the method gives only bounds for higher numbers. For example, using this technique one can show that $38 \le R(5, 5)$, but in fact other techniques have shown that $43 \le R(5, 5)$. We present all of the known nontrivial Ramsey numbers in Table 4.1. The notation l/u means that l and u are the best known lower and upper bounds for that particular Ramsey number. We refer readers to [11] and to the dynamic survey by S. Radziszowski found in the *Electronic Journal of Combinatorics* at http://www.combinatorics.org.

Table 4.1 Ramsey numbers $R(a, b)$.

a \ b	3	4	5	6	7	8	9
3	6	9	14	18	23	28	36
4		18	25	36/41	49/61	56/84	73/115
5			43/49	58/87	80/143	101/216	126/316
6				102/165	113/298	132/495	169/780
7					205/540	217/1031	241/1713
8						282/1870	317/3583
9							565/6588

Open problem. Determine $R(5, 5)$.

Open problem. Determine a formula for $R(n, n)$.

We know that $R(5, 5) \le R(4, 5) + R(5, 4) = 50$. Unfortunately, this still leaves an enormous computation problem in evaluating $R(5, 5)$. The naive approach, examining all $2^{\binom{49}{2}}$ labeled graphs on 49 vertices, is intractable.

When we consider more than two colors, the only known nontrivial Ramsey number is $R(3, 3, 3) = 17$, whose proof we leave as an exercise. The only known nontrivial t-uniform hypergraph Ramsey number with $t \ge 3$ is $R(4, 4; 3) = 13$. This state of limited knowledge is exasperating because Ramsey

numbers are intimately connected with other numbers and functions, as we shall see later in this chapter. Any new Ramsey number would be very valuable.

Let us now consider lower and upper bounds for the diagonal Ramsey numbers $R(a, a)$. The trivial lower bound $R(a, a) > (a-1)^2$ is immediate: join with green edges $a-1$ disjoint copies of a red K_{a-1}; this coloring has no monochromatic K_a. A more sophisticated lower bound is obtained by the probabilistic method in the next section.

To find an upper bound we use Pascal's identity. We recall that $R(a, 2) = a$ for $a \geq 2$ and $R(a, b) \leq R(a-1, b) + R(a, b-1)$ for $a, b \geq 3$.

Upper bound for Ramsey numbers. For all $a, b \geq 2$, we have $R(a, b) \leq \binom{a+b-2}{a-1}$.

Proof We use induction on a and b, noting that $R(a, 2) = a = \binom{a}{a-1}$ and $R(2, b) = b = \binom{b}{b-1}$, so the inequality holds when $b = 2$ or $a = 2$. Suppose that the inequality holds for $R(a-1, b)$ and $R(a, b-1)$ for $a, b \geq 3$. Then

$$\begin{aligned} R(a,b) &\leq R(a-1,b) + R(a,b-1) \\ &\leq \binom{a+b-3}{a-2} + \binom{a+b-3}{a-1} \\ &= \binom{a+b-2}{a-1}, \end{aligned}$$

and the inequality is established.?

For diagonal Ramsey numbers, the upper bound becomes $R(a, a) \leq \binom{2a-2}{a-1}$, and we can determine a asymptotic estimate. One of the great open problems of Ramsey theory is to calculate $\lim_{a\to\infty} R(a, a)^{1/a}$ (if it exists).

It follows that

$$\begin{aligned} R(a,a) &\leq \binom{2a-2}{a-1} \\ &< 2^{2a-2} \\ &= 4^{a-1}. \end{aligned} \tag{4.9}$$

Thus, we obtain an asymptotic upper bound for $R(a, a)^{1/a}$:

$$\limsup R(a,a)^{1/a} \leq \limsup 4^{(a-1)/a} = 4. \tag{4.10}$$

Using Stirling's asymptotic approximation to the factorial function,

$$n! \sim n^n e^{-n} (2\pi n)^{1/2}, \tag{4.11}$$

we can improve the upper bound a little. Since $\binom{2a-2}{a-1} < \binom{2a}{a}$, and

$$\binom{2a}{a} \sim \frac{(2a)^{2a} e^{-2a} (2\pi 2a)^{1/2}}{a^{2a} e^{-2a} 2\pi a}, \tag{4.12}$$

it follows that

$$R(a,a) < \frac{4^a}{(\pi a)^{1/2}} (1 + o(1)),$$

where $o(1)$ is a function of a which tends to 0 as a tends to ∞.

A lower bound for $\liminf R(a, a)^{1/a}$ is determined in the next section.

EXERCISES

4.11 Find a 2-coloring of K_8 that proves $R(3, 4) > 8$.

4.12 Find a 2-coloring of K_{13} that proves $R(3, 5) > 13$.

4.13 Prove $R(3, 3, 3) = 17$.

4.14 Prove that if K_{327} is 5-colored, there exists a monochromatic K_3.

4.4 The probabilistic method

To obtain a good lower bound for $R(a, a)$, we turn to the probabilistic method, a technique used widely throughout existential combinatorics. See [2]. The idea is to turn the objects in question (green–red colorings of a graph) into events in a probability space and demonstrate that a desired event (a coloring containing no monochromatic subgraph of specified size) occurs with positive probability. If D is a set of desired objects in a sample space S, then the probability that a random object is desired equals $\Pr(D) = |D|/|S|$. If we can show that $\Pr(D)$ is positive, then D is nonempty and there *exists* a desired object. Usually, probabilistic arguments can be framed directly in terms of the cardinalities $|D|$ and $|S|$. However, the probabilistic language has undisputed bookkeeping and conceptual advantages in proving complex theorems. To illustrate the distinction and parallelism between the two points of view, we present two proofs of the following lower bound for $R(a, a)$, one in terms of cardinality and the other in terms of probability.

Lower bound for Ramsey numbers. If $\binom{n}{a}2^{1-\binom{a}{2}} < 1$, then $R(a, a) > n$.

Proof 1 (Cardinality). Because each of the $\binom{n}{2}$ edges of K_n may be colored independently, the number of green–red colorings is $2^{\binom{n}{2}}$. The number of green–red colorings of K_n, with a monochromatic K_a is $|\cup A_S|$, where A_S is the collection of green–red colorings in which the subgraph S is monochromatic and S ranges over all possible subgraphs of K_n isomorphic to K_a. We bound $|\cup A_S|$ as follows:

$$\begin{aligned}\left|\bigcup A_S\right| &\le \sum_S |A_S| \\ &= 2 \cdot \binom{n}{a} 2^{\binom{n}{2}-\binom{a}{2}} \\ &< 2^{\binom{n}{2}}.\end{aligned}$$

The first inequality is an enumeration estimate proved by induction on the number of terms in the union; it also follows from the inclusion–exclusion principle. The equality follows from the observation that there are $\binom{a}{2}$ copies of K_a inside K_n. Since each K_a is monochromatic, there are two choices for the color of its edges. The remaining $\binom{n}{2} - \binom{a}{2}$ edges of K_n are colored green or red arbitrarily.

Now, because $|\cup A_S|$ is less than the total number of green–red colorings of K_n, there is a coloring which does not contain a monochromatic K_a. Therefore $R(a, a) > n$.

Proof 2 (Probability). Suppose the edges of K_n, are randomly and independently colored green or

red. Think of flipping a coin for each edge. If the coin lands heads, then the edge is colored green; if it lands tails, then red. For each subgraph S of K_n isomorphic to K_a, let A_S be the event that S is monochromatic. We have

$$\begin{aligned}\Pr(A_S) &= \Pr(S \text{ is green}) + \Pr(S \text{ is red}) \\ &= 2^{-\binom{a}{2}} + 2^{-\binom{a}{2}} \\ &= 2^{1-\binom{a}{2}}.\end{aligned}$$

Therefore

$$\begin{aligned}\Pr\left(\bigcup A_S\right) &\le \sum_S \Pr(A_S) \quad \text{(subadditivity of probability)} \\ &= \binom{n}{a} 2^{1-\binom{a}{2}} \\ &< 1.\end{aligned}$$

The complement of the event $\cup\, A_S$ occurs with positive probability; hence there exists a desired configuration–a 2-coloring of K_n, with no monochromatic K_a. Again, $R(a, a) > n$.

The theorem contains an implicit lower bound for $R(a, a)$, if we can untangle it. Fix a and let N be the minimum value of n satisfying $\binom{n}{a} 2^{1-\binom{a}{2}} \ge 1$. Then

$$\begin{aligned}R(a,a) &\ge N \\ &= (N^a)^{1/a} \\ &> \left[\binom{N}{a} a!\right]^{1/a} \\ &\ge \left(2^{\binom{a}{2}-1} a!\right)^{1/a} \\ &= 2^{a/2-1/2-1/a} a!^{1/a}. \qquad (4.13)\end{aligned}$$

From Stirling's asymptotic formula for the factorial function, it follows that

$$R(a,a) > a 2^{a/2} \left[\frac{1}{e\sqrt{2}} + o(1)\right]. \qquad (4.14)$$

Finally, we have

$$\liminf R(a,a)^{1/a} \ge \liminf \left\{a 2^{a/2} \left[\frac{1}{e\sqrt{2}} + o(1)\right]\right\}^{1/a} = \sqrt{2}. \qquad (4.15)$$

Combining this lower bound with our previously obtained upper bound, we obtain bounds on $\lim_{a\to\infty} R(a, a)^{1/a}$ (if it exists):

$$\sqrt{2} \le \liminf R(a,a)^{1/a} \le \lim R(a,a)^{1/a} \le \limsup R(a,a)^{1/a} \le 4. \qquad (4.16)$$

These are the best bounds known at present.

Open problem Determine whether $\lim_{a\to\infty} R(a, a)^{1/a}$ exists and find its value.

In 1995 J. H. Kim proved the first conclusive result about the growth of $R(n, k)$ for fixed k. He showed that the order of magnitude of $R(n, 3)$ is $n^2/\log n$. See [6].

Open problem. Determine the asymptotic behavior of $R(n, 4)$.

EXERCISES

4.15 Obtain a lower bound for $R(100, 100)$.

4.16 Use the probabilistic method to prove that almost all labeled graphs have diameter 2; hence almost all labeled graphs are connected.

4.17 Use the probabilistic method to prove Schütte's theorem For every m there exists a tournament T such that for each $S \subseteq T$, $|S| = m$, there exists a vertex $p \in T - S$ which is directed to each vertex of S. Find such tournaments for $m = 1$ and $m = 2$.

Hint: The tournament for $m = 2$ can be constructed from the set of quadratic residues modulo 7 as follows. Let R_p and N_p be the set of quadratic residues and nonresidues modulo 7, respectively. Put a directed arrow from vertex i to vertex j if $i - j \in R_p$ and an arrow from j to i if $i - j \in N_p$. Check that this tournament has the desired property.

Also prove that this tournament is unique up to isomorphism.

Hint: First prove that every vertex has outdegree 3. Next prove that if vertex a is directed to vertices b, c, d, then b, c, d form a cyclic triple.

Schütte's theorem was proved by Paul Erdős in 1963.

4.5 Schur's theorem

In this section, we prove a proposition about equations as a corollary of Ramsey's theorem, and in the next section we prove van der Waerden's theorem, an elegant statement about arithmetic progressions. The theme is that of finding order in disorder.

Given $c, n \geq 1$, we consider functions $f : \mathrm{N}_n \to \{A_1,\ldots,A_c\}$. As usual, we think of the A_i, as colors and f as assigning a color to each integer, thereby partitioning N_n into color classes. If S is a set of positive integers and f restricted to S is a constant function, then S is *monochromatic*. What kinds of monochromatic sets can we find given that n is large enough compared to c? One answer to this question was provided by Issai Schur (1875–1941) in 1916.

Schur's theorem. For each $c \geq 1$, there exists a least integer $n = S(c)$ with the following property: For any function $f : \mathrm{N}_n \to \{A_1,\ldots,A_c\}$, there exists an A_i, containing x, y, z (with $x = y$ allowed) such that $x + y = z$. In other words, there is a monochromatic solution to the equation $x + y = z$.

Proof. Let $m = R(3,\ldots,3) - 1$, where $\mathrm{R}(3,\ldots,3)$ is the c-color Ramsey number that guarantees a monochromatic triangle. We claim that m has the desired property, and hence $S(c)$ exists and satisfies $S(c) \leq m$. The function $f : \mathrm{N}_m \to \{A_1,\ldots,A_c\}$ generates a c-coloring of the complete graph on vertices $1, 2,\ldots,m + 1$ by assigning to edge ij the color that has been assigned to the integer $|i - j|$. The presence of a monochromatic triangle on, say, vertices a, b, c $(a < b < c)$ implies that the equation $x + y = z$ has the monochromatic solution $(b - a) + (c - b) = (c - a)$.

Although it is considered an important part of Ramsey theory, Schur's theorem was introduced by Schur in an attempt to prove Fermat's last theorem. See Notes.

The integers $S(c)$ are called c-color *Schur numbers*. It is trivial to observe that $S(1) = 2$ ($1 + 1 = 2$). We leave it to the reader to check that $S(2) = 5$ and $S(3) = 14$. The only other known Schur number is $S(4) = 45$. Thus there is a general state of ignorance about Schur numbers, although they are linked to the equally mysterious Ramsey numbers by the inequality

(4.17) $S(c) \leq R(3, \ldots, 3) - 1.$

Open problem. Find the value of $S(5)$.

EXERCISES

4.18 Prove $S(2) = 5$ and $S(3) - 14$.

4.19 Suppose that we have a sum-free c-coloring of $\{1,\ldots,n\}$, with the partition $\{A_1,\ldots,A_c\}$. Then we can obtain a sum-free $(c + 1)$-coloring of $\{1,\ldots,3n + 1\}$ as follows. For each i and each $a \in A_i$ include in A_i the new element $a + 2n + 1$. Define $A_{c+1} = \{n + 1,\ldots,2n + 1\}$. Show that this procedure works. What bounds does it give on $S(c)$ for various values of c and in general?

4.20 Let $f(n)$ be the minimum number of triples (x, y, z) such that $x + y = z$ and $x \neq y$ when $\{1, 2, \ldots,n\}$ is 2-colored. Conjecture a formula for $f(n)$. Such a formula was found by T. Schoen.

4.6 Van der Waerden's theorem

Schur's theorem states that any coloring function $f : N_n \rightarrow \{A_1,\ldots,A_c\}$ forces a monochromatic solution to the equation $x + y = z$ (whenever n is sufficiently large compared to c). What other monochromatic structures are forced? One direction for generalization is provided by Rado's theorem, which asserts the existence of a monochromatic solution to the equation

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

as long as some nonempty subset of the α_i, sums to 0. If this condition is met, we say that the equation is *regular*. For example, the equation $2x_1 - 7x_2 + 3x_3 + 4x_4 - 6x_5 = 0$ is regular ($-7 + 3 + 4 = 0$). Another direction is provided by B. L. van der Waerden's 1927 theorem concerning arithmetic progressions. An *arithmetic progression* of length l (or l-AP) is a sequence

$$a, a + d, a + 2d, \ldots, a + (l - 1)d$$

of l numbers (integers, say), where each consecutive pair differ by a constant number $d \geq 1$. For example, the sequence 20, 30, 40, 50, 60, 70 is a 6-AP. Van der Waerden's theorem asserts the existence of a monochromatic l-AP when N_n, is partitioned into c classes (and n is sufficiently large with respect to c and l).

Van der Waerden's theorem (1927). Given $c \geq 1$ and $l \geq 1$, there exists a least integer $W = W(c, l)$ with the following property: If N_W is partitioned into c classes $A_1,\ldots,A_c$, then one of the classes A_i contains a monochromatic l-AP.

Proof. The proof is by induction on c and l. In this proof, we use the notation $[n]$ for N_n. The theorem is trivially true for some ordered pairs c, l, and in these cases we can actually determine the values of $W(c, l)$: $W(1, l) = l$, $W(c, 1) = 1$, $W(c, 2) = c+ 1$. The first and third of these statements are the basis of the induction. We shall assume the existence of $W(d, l)$ for every d and prove the existence of $W(c, l + 1)$. The reader is encouraged to envision a table of c and l, and judge whether this plan would really cover all ordered pairs c, l. We claim that $W(c, l + 1)$ exists and satisfies $W(c, l + 1) \leq f(c)$, where f is defined recursively:

$$\begin{aligned} f(1) &= 2W(c, l) \\ f(n) &= 2W(c^{f(n-1)}, l)f(n-1), \quad n \geq 2. \end{aligned} \tag{4.18}$$

As in the proof of Ramsey's theorem, we are establishing an existence result by constructing an upper bound. However, the formulas in the upper bound grow too rapidly to furnish much insight into the exact values of $W(c, l)$.

Suppose that $[f(c)]$, which we call a *c-block*, is c-colored without a monochromatic $l + 1$-AP, and $[f(c)]$ is partitioned into $f(c)/f(c - 1)$ blocks of $f(c - 1)$ consecutive integers, which we call $(c - 1)$-*blocks*. Likewise, each $(c - 1)$-block is partitioned into $f(c - 1)/f(c - 2)$ blocks of $f(c - 2)$ consecutive integers, which we call $(c - 2)$-*blocks*. This partitioning happens at each of the c levels, until, at last, each 1-block is partitioned into $2W(c, l)$ 0-*blocks* (which are just integers).

By definition of $W(c, l)$, the first half of each 1-block contains a monochromatic l-AR Here occurs the first leap of inspiration in the proof. The coloring of the elements of a 1-block *induces* a coloring of the 1-block itself. That is, we assign one of $c^{f(1)}$ colors to the 1-block according to the way its elements are c-colored. Because $f(2) = 2W(c^{f(1)}, l)f(1)$, each 2-block contains $2W(c^{f(1)}, l)$ 1-blocks, so that, by definition of $W(c^{f(1)}, l)$, the first half of each 2-block contains a monochromatic l-AP of 1-blocks. Similarly, the first half of each 3-block contains a monochromatic l-AP of 2-blocks. This construction happens at each level, so that the first half of $[f(c + 1)]$ contains a monochromatic l-AP of c-blocks. Let us consider only those integers which lie in l-APs at *all* c levels of blocks. We coordinatize each integer as

$$x = (x_1, \ldots, x_c),$$

with $1 \leq x_i \leq l$, where x_i is the position of x in the monochromatic l-AP of the i-block in which it resides. All coordinatized integers have the same color, say A_1. Within each 1-block, the l integers

$$(1, x_2, \ldots, x_c), (2, x_2, \ldots, x_c), \ldots, (l, x_2, \ldots, x_c)$$

constitute a monochromatic l-AP. Therefore, the integer $(l + 1, x_2,\ldots,x_c)$ has a color *other than* A_1, say A_2. Furthermore, the factor 2 in the definition of $f(1)$ implies that $(l + 1, x_2,\ldots,x_c)$ occurs within the 1-block. (The 2 is a convenient constant used to stretch the block enough to accommodate the $(l + 1)$st term of an AP.) Here occurs the second leap of inspiration: the idea of *focusing*. Within a 2-block, the l integers

$$(l+1, 1, x_3, \ldots, x_c), (l+1, 2, x_3, \ldots, x_c), \ldots, (l+1, l+1, x_3, \ldots, x_c)$$

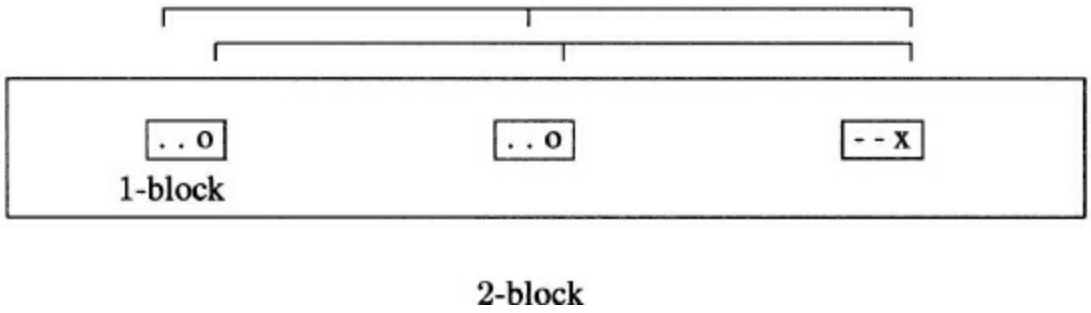
are a monochromatic l-AP of color A_2. This forces $(l + 1,l + 1, x_3,\ldots,x_c)$ to be a color other than A_2. However, we can focus a second l-AP on this integer, namely,

$$(1, 1, x_3, \ldots, x_c), (2, 2, x_3, \ldots, x_c), \ldots, (l, l, x_3, \ldots, x_c).$$

Thus, $(l + 1,l + 1, x_3,\ldots,x_c)$ cannot be color A_1 or A_2; say it is colored A_3. Figure 4.3 illustrates the

two focused progressions, representing colors A_1, A_2, A_3 by dots, circles, and an x, respectively. (The dashes represent numbers with undetermined colors.) Continuing this focusing process at each of the c levels, we conclude that $(l + 1, l + 1, \ldots, l + 1)$ can be none of the colors $A_1, \ldots, A_c$, a contradiction. Therefore, there exists a monochromatic $(l + 1)$-AP.

Figure 4.3 Two 3-APs focusing on an integer.



The values of $W(c, l)$ are called *Van der Waerden numbers*. As we remarked in the proof, the inequality $W(c, l + 1) \leq f(c)$ does little to establish good estimates for them. In fact, the state of knowledge is even worse for van der Waerden numbers than for Ramsey numbers. The seven known nontrivial van der Waerden numbers are listed in Table 4.2. The proof of one of these values, $W(2, 3) = 9$, is called for in the exercises.

Table 4.2 The known van der Waerden numbers $W(c, l)$ with $c \geq 2$, $l \geq 3$.

c \ l	3	4	5	6
2	9	35	178	1132
3	27	293		
4	76			

Open problem. Find $W(5, 3)$.

Although van der Waerden's theorem asserts the existence of a monochromatic l-AP, it does not tell us *which* color it is. The following theorem, whose proof is beyond the scope of this book, guarantees the existence of a monochromatic l-AP in any color that occurs "with positive density." We define the *density function* of a set S of positive integers to be

$$d(S, n) = \frac{|\mathbf{N}_n \cap S|}{n}. \tag{4.19}$$

The density function measures the fraction of the first n integers which occur in S. Clearly, $0 \leq d(S, n) \leq 1$ for all S and n.

Szemerédi's theorem. For all real numbers $d > 0$ and all $l \geq 1$, there is a positive integer $N(d, l)$ with the following property: If $n \geq N(d, l)$ and $S \subseteq \{1, \ldots, n\}$ with $d(S, n) \geq d$, then S contains an l-AP.

Paul Erdős (1913–1996) conjectured the above result in 1935, but it was not proved until 1975 by Endre Szemerédi. In 1977 Hillel Furstenberg gave a proof using ergodic theory.

Conjecture. (Erdős) If $\{a_i\} \subseteq \mathbf{N}$ **and** $\sum \frac{1}{a_i}$ is a divergent series, then $\{a_i\}$ contains arbitrarily long arithmetic progressions.

It is well known that $\Sigma\ 1/p_i$ diverges if $\{p_i\}$ is the set of primes (see [15]), and in 2006 Ben Green and Terence Tao proved that there exist arbitrarily long arithmetic progressions of primes. Erdős' conjecture is still open. In 2010 a 26-AP of primes was found:

$$43{,}142{,}746{,}595{,}714{,}191 + 5{,}283{,}234{,}035{,}979{,}900n, \quad 0 \le n \le 25.$$

EXERCISES

4.21 Prove $W(2, 3) = 9$.

4.22 Find upper bounds for $W(3, 4)$ and $W(4, 4)$.

4.23 Prove or disprove: If N is 2-colored, then there exists a monochromatic infinite AP.

4.24 Prove or disprove: If R is 2-colored, then there exist $a, b, c \in \mathbf{R}$ with a, b, c all the same color and $(c - b)/(b - a) = \sqrt{2}$.

4.25 (Putnam Competition, 1960) Consider the arithmetic progression

$$a, a + d, a + 2d, \ldots,$$

where a and d are positive integers. For any positive integer k, prove that the progression has either no exact kth powers or infinitely many.

4.26 Find a 6-AP of prime numbers.

4.27 Prove that for any positive integers c and l, there exists a number W with the property that, whenever the set N_W is c-colored, there exists an l-AP with each of its terms and the common difference the same color.

4.28 Using the compactness principle, prove that the following theorem is equivalent to van der Waerden's theorem: For all $c, l \ge 1$, no matter how N is c-colored, there exists a monochromatic l-AP.

4.29 With the notation of Szemerédi's theorem, suppose that there exists a density $d < 1$ such that $N(d, l)$ exists for all $l \ge 1$. Prove that $N(d^2, l)$ exists by showing that it satisfies

$$N(d^2, l) \le N(d, l) \cdot N(d, W(N(d, l), l)),$$

where W is the van der Waerden function. Thus conclude that $N(d, l)$ exists for arbitrarily small $d > 0$ and all $l \ge 1$.

4.30 Let $r_k(n)$ be the greatest integer l such that there is a sequence of integers $1 \le a_1 < \cdots a_l \le n$ which does not contain an l-AP. Prove that

$$r_k(m + n) \le r_k(m) + r_k(n).$$

Prove that this implies that

$$\lim_{n \to \infty} \frac{r_k(n)}{n}$$

exists for each k.

Notes

For original papers of Frank Ramsey, Paul Erdős and George Szekeres, and R. P. Dilworth, see [9].

Ramsey numbers have been generalized in many ways. For example, in 1972 Václav Chvátal and Frank Harary defined the *graph Ramsey number r(G, H)* to be the minimum number of vertices in a complete graph which, when 2-colored, yields a green subgraph G or a red subgraph H. They showed that

$$r(G, H) > (\chi(G) - 1)(p(H) - 1),$$

where $\chi(G)$ is the chromatic number of G and $p(H)$ is the number of vertices of H. They used this inequality to prove $r(T_m, K_n) = (m - 1)(n - 1) + 1$, where T_m is a tree with m vertices. See [11].

Schur's theorem was proven by Issai Schur in an attempt to prove Fermat's last theorem (FLT). Although Schur didn't prove FLT, he did prove that, for all n, if p is prime and sufficiently large, then the congruence $x^n + y^n \equiv z^n$ has a nonzero solution modulo p. Briefly, the argument is to suppose that p is prime and greater than $S(n)$. Thus if $\{1,\ldots,p - 1\}$ is n-colored, there exists a monochromatic subset $\{a, b, c\}$ with $a + b = c$. Let $H = \{x^n : x \in Z^*_p\}$, a subgroup of Z^*_p of index $\gcd(n, p - 1) \leq n$. The cosets of Z^*_p define an n-coloring f of Z^*_p such that $f(a) = f(b) = f(c)$ and $a + b = c$. This implies that $1 + a^{-1}b = a^{-1}c =$ (in Z_p), and in fact 1, $a^{-1}b$, and $a^{-1}c$ are all nth powers in Z_p.

B. L. van der Waerden (1903–1996) proved his 1927 theorem as a generalization of the following conjecture of Schur: If **N** is partitioned into two classes, then one of the classes contains arbitrarily long arithmetic progressions.

Ramsey's theorem (in its various formulations) and van der Waerden's theorem are usually thought of as the two cornerstone theorems of Ramsey theory. See [11] for a further discussion of these theorems and other theorems of Ramsey theory, including Gallai's theorem, Rado's theorem, Folkman's theorem, and the Hales–Jewett theorem.

CHAPTER 5

ERROR-CORRECTING CODES

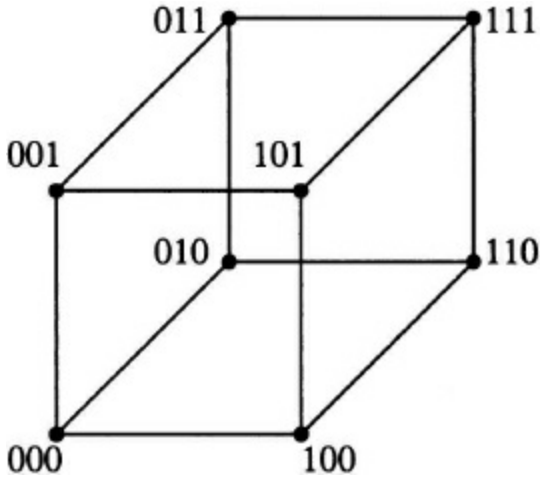
Sixteen unit hyperspheres can be arranged in $\mathbf{R}^7$ so that each hypersphere is tangent to exactly seven of the other hyperspheres.

This configuration of spheres is called a *perfect packing*. How do we obtain such a packing? What makes it perfect? What are its combinatorial properties? In this chapter and the next, we investigate such combinatorial designs, paying close attention to the interrelationships among the constructions and often finding equivalences between seemingly different structures. As a capstone, we construct the $(23, 2^{12}, 7)$ Golay code G_{23}, the $S(5, 8, 24)$ Steiner system, and Leech's 24-dimensional lattice L. We begin our tour of combinatorial constructions with practical examples called codes.

5.1 Binary codes

Let $F = \{0, 1\}$, the field of two elements. Then F^n, the collection of strings of length n over F, is a vector space of dimension n over F. We can picture F^n as the set of vertices of the n-dimensional unit hypercube. For example, Figure 5.1 depicts F^3 as the set of vertices of the cube. These vertices are coordinatized with the eight vector representatives 000, 001, 010, 011, 100, 101, 110, and 111. Note that two vertices are edge adjacent if and only if their vector representatives differ in exactly one coordinate.

Figure 5.1 F^3 as the set of vertices of a cube.



Given any two binary strings $v, w \in F^n$, we define the *Hamming distance* $d(v, w)$ between v and w to be the number of coordinates where v and w differ. This is also the shortest edge path in the hypercube between x and y. For example, we can see in Figure 5.1 that $d(011, 101) = 2$.

The function d is a metric, which we call the *Hamming metric* or *Hamming distance*, and F^n is a metric space.

Theorem. The function d is a metric on F^n; that is, for all $v, w, x \in F^n$, the following properties hold:

(1) $d(v, w) \geq 0$ with equality only when $v = w$ (positivity);

(2) $d(w, v) = d(v, w)$ (symmetry);

(3) $d(v, w) + d(w, x) \geq d(v, x)$ (triangle inequality).

Proof. Properties (1) and (2) are immediate from the definition of d. We prove the triangle inequality by verifying that it is preserved componentwise. Let v_i, w_i, x_i be the ith components of the vectors v, w, x, respectively. If $v_i = w_i = x_i$, then the contribution to both sides of the inequality is 0. If not, then the contribution to the left side is at least 1 while the contribution to the right side is at most 1. Hence, the inequality is preserved componentwise. (It is also easy to "see" that d is a metric by realizing that $d(x, y)$ is the shortest path length between x and y along the edges of a cube embedded in $\mathbf{R}^n$.)

The *weight* $w(v)$ of a vector v is the number of 1′s in the vector representation of v. A simple componentwise proof demonstrates that $d(x, y) = w(x - y)$ for any $x, y \in F^n$.

The *Hamming sphere* with radius r and center c is the set of all $v \in F^n$ such that $d(v, c) \leq r$. The volume of the sphere is the number of elements in it: $\sum_{k=0}^{r} \binom{n}{k}$. Note that $\binom{n}{k}$ counts selections of the k coordinates in which c and v disagree.

It is difficult to picture spheres when n is large (and they don't look very spherical). We will mainly be interested in how densely they can be packed, because, as we shall see, dense packings signify good codes.

A *code* A is a subset of F^n with $|A| \geq 2$. The elements of A are called *codewords*. In real-life applications, information can be sent reliably over a noisy channel by encoding redundancy in the message. A codeword $v \in A$ is transmitted and a possibly distorted vector v' is received. As it might happen that v' equals a codeword in A different from v, it is not always possible to tell whether any errors have been committed in the transmission. However, if the Hamming distances between pairs of codewords are fairly large (which is achievable with redundancy), it is unlikely that v' will equal another codeword. If the Hamming distances are large enough, it may be possible to detect errors when they occur and correct them.

The *distance* $d(A)$ of a code A is the minimum Hamming distance between distinct codewords in A. For example, the code $A = \{011, 101, 110\} \subseteq F^3$ has distance $d(A) = 2$, because any two of the vectors in A differ by two bits. It is always true that $1 \leq d(A) \leq n$.

A code with distance $d = e + 1$ *detects* e errors. A code with distance $d = 2e + 1$ *corrects* e errors. To justify these definitions, suppose that a codeword v is sent and a string v' is received, with $1 \leq d(v, v') \leq e$. If $d = e + 1$, then v' cannot equal some codeword x or else $d(v, v') \geq e + 1$, a contradiction. Therefore, we can detect that at least one error has occurred. If $d = 2e + 1$, then v' cannot have resulted from the transmission of an erroneous codeword x (and at most e errors), or else $2e + 1 \leq d(v, x) \leq d(v, v') + d(v', x) \leq e + e = 2e$, a contradiction. Therefore, we can identify the particular vector v that was sent and correct the errors.

EXAMPLE 5.1 Triplicate code

The code $\{000, 111\} \subseteq F^3$ is called a *triplicate code*. Under the map $0 \mapsto 000$ and $1 \mapsto 111$, each bit is tripled. If an error occurs (a bit is switched from 0 to 1 or 1 to 0), then we can still tell which message was intended. Thus, the code has distance 3 and is capable of correcting one error. If a codeword is transmitted and we receive 001, then, under the assumption that at most one error has been committed, the intended word must be 000. Let q be the probability that a bit is mistakenly altered from 0 to 1 or from 1 to 0, and suppose that bits are altered independently of each another. The decoding scheme fails if two or three errors occur, and this happens with probability $3q^2(1 - q) + q^3$, which is asymptotic to $3q^2$ for q small. This probability of failure is much smaller than the probability q of failure when no code is used. However, the increase in reliability is paid for by a decrease in transmission rate. The *rate* $r(A)$ of a code A, defined as $r(A) = \log_2 |A|/n$, measures the amount of information per code bit conveyed over the communication channel. Always, $1/n \leq r(A) \leq 1$. In the above example, $r(A) = -(\log_2 2)/3 = 1/3$, which means that when information is sent in triplicate the rate decreases by a factor of 3 (which is reasonable).

We refer to a code $A \subseteq F^n$ with distance $d(A) = d$ as an $(n, |A|, d)$ code. The number n is sometimes called the *dimension* of the code, and $|A|$ is called the *size* of the code. Let there be no confusion between the distance d (an integer) and the Hamming metric d (a function).

For n fixed, there is an inverse relationship between $|A|$ and d (and therefore between the rate and the error-correcting capability of the code). The *fundamental problem of coding theory* is to find codes with high rate and large distance. The next theorem gives sharp focus to this problem.

Hamming upper bound. If A corrects e errors, then

$$|A| \leq \frac{2^n}{\sum_{k=0}^{e} \binom{n}{k}}. \tag{5.1}$$

Proof. Because $d \geq 2e + 1$, the spheres of radius e centered at the codewords do not intersect. Therefore, the total volume of the spheres is at most the cardinality of F^n; that is,

$$|A| \sum_{k=0}^{e} \binom{n}{k} \leq 2^n,$$

from which the upper bound follows instantly.

The Hamming upper bound for $|A|$ is also called the *sphere packing bound.*

EXERCISES

5.1 Prove that $d(x, y) = w(x - y)$ for any $x, y \in F^n$.

5.2 Prove that $d(x, y) = d(x + z, y + z)$ for any $x, y, z \in F^n$.

5.3 Prove that $w(x + y) = w(x) + w(y) - 2x \cdot y$ for any $x, y \in F^n$.

5.4 Find a code in F^4 with eight codewords and Hamming distance 2. How many errors can this code detect?

5.2 Perfect codes

If the Hamming upper bound is achieved for a code A, we say that A is *perfect*. Perfect codes correspond to sphere packings of F^n with no wasted space (vectors not in a sphere). If A is perfect, then $\sum_{k=0}^{e} \binom{n}{k}$ must divide 2^n and hence be a power of 2. We will soon see that this rarely happens.

If $|A| = 2$, then the maximum value of $d(A)$ is n and this value is achieved, for instance, when A consists of the all-0 vector and the all-1 vector. If $|A| > 2$, then some two vectors must agree on any given component, so $d(A) < n$. Therefore, $n > d \geq 2e + 1$, which implies that

$$e < \frac{n-1}{2}. \tag{5.2}$$

If $e = 1$, then $\sum_{k=0}^{e} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} = 1 + n$, which is a power of 2 when $n = 2^r - 1$ for some $r \geq 2$. In this case, $|A| = 2^n/2^r = 2^{n-r} = 2^m$, where $m = n - r = 2^r - 1 - r$. Hence, the parameters of such a code are $(n, |A|, d)$ $(2^r - 1, 2^m, 3)$.

The next theorem says that there are only two feasible sets of parameters for perfect codes when $1 < e < (n-1)/2$. This was proved by Aimo Tietäväinen and J. H. van Lint (1932–2004). Unfortunately, no simple proof of this fact is known. Their proof uses the theory of equations and is quite complicated. See [22].

Theorem. The only values of n and e for which $1 < e < \frac{n-1}{2}$ and $\sum_{k=0}^{e} \binom{n}{k}$ is a power of 2 are $(n, e) = (23, 3)$ and $(90, 2)$.

These values correspond to two special sums of binomial coefficients:

$$1 + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 2^{11}$$

$$1 + \binom{90}{1} + \binom{90}{2} = 2^{12}.$$

If there are codes with these values of n and e, they would have parameters $(n, |A|, d) = (23, 2^{12}, 7)$ and $(90, 2^{78}, 5)$. We outline a proof in the exercises that there is no code with parameters $(90, 2^{78}, 5)$. A code with parameters $(23, 2^{12}, 7)$, called the Golay code G_{23}, is constructed in Section 6.7.

We have taken the base field of our codes to be GF(2), but if we allow other base fields, it turns out that there is only one more perfect code, the ternary code G_{11} with parameters $(11, 3^6, 5)$, discovered by Marcel Golay. If we consider any alphabet as the base set (not necessarily a field), then a perfect code is one of the two Golay codes G_{23} and G_{11} or else has the parameters of a Hamming code.

In the next section, we will describe the Hamming codes, a family of perfect 1-error-correcting codes.

EXERCISES

5.5 Prove that the maximum number $A(n, e)$ of codewords in an e-error-correcting code in F_n satisfies the *Gilbert lower bound*

$$A(n,e) \geq \frac{2^n}{\sum_{k=0}^{2e} \binom{n}{k}}.$$

5.6 Show that there exists no code $A \subseteq \mathrm{GF}(2)^{10}$ with 19 words and $d(A) = 5$.

5.7 Use a computer to verify that the only ordered pairs (n, e) with $2 \leq e < (n-1)/2 \leq 50$ and $\sum_{k=0}^{e} \binom{n}{k}$ a power of 2 are (23, 3) and (90, 2).

5.8 Prove that there is no $(90, 2^{78}, 5)$ code.

Hint: Suppose that there is such a code. Without loss of generality, we can assume that the code contains the zero vector (why?). The code, being perfect, corresponds to a sphere packing of F^{90} with 2^{78} spheres of radius 2. Let X be the set of weight 3 vectors in F^{90} which have 1′s in the first two components. Show that X has 88 elements. How are the elements of X partitioned by spheres around codewords of weight 5?

5.9 Show that $(11, 3^6, 5)$ are feasible parameters for a perfect ternary code.

5.3 Hamming codes

As in the previous section, we define $n = 2^r - 1$ and $m = 2^r - 1 - r$, where $r \geq 2$. If $r = 2$, then $n = 3$ and $m = 1$, and the vertices 000 and 111 of the cube C^3 of Figure 5.1 constitute such a code. We now describe a code $A \subseteq F^n$ with $|A| = 2^m$ that corrects one error, exhibiting a construction in the case $r = 3$, $n = 7$, $m = 4$. The constructions for $r > 3$ are carried out similarly.

Let

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{5.3}$$

be the $r \times n = 3 \times 7$ matrix whose columns are the numbers $1,\ldots,n$ written in binary. The matrix H represents a linear transformation from F^7 to F^3:

$$H: F^7 \longrightarrow F^3$$
$$v \longmapsto Hv.$$

We define the code A to be the kernel of H; that is,

$$A = \{v \in F^7 : Hv = 0\}. \tag{5.4}$$

(The 0 here is the 3×1 zero vector.) We call H a *parity check matrix* for A. Any code A for which $x \in A$ and $y \in A$ imply $x + y \in A$ is called a *linear* code. Clearly, a code described as the kernel of a parity check matrix is a linear code.

By inspection, we find two of the vectors belonging to A:

$$\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}1\\1\\1\\0\\0\\0\\0\end{bmatrix}.$$

As the Hamming distance between these two vectors is 3, we see that $d(A) \leq 3$. We need to prove that $d(A) \geq 3$ and $|A| = 2^m = 16$.

Assume that $v = [x\ y\ a\ z\ b\ c\ d]^t \in A$. (The reason for the nonalphabetical listing of the components of v will become clear in a moment.) Because v is in the kernel of H, we have $Hv = 0$. Thus

$$\begin{bmatrix}1&0&1&0&1&0&1\\0&1&1&0&0&1&1\\0&0&0&1&1&1&1\end{bmatrix}\begin{bmatrix}x\\y\\a\\z\\b\\c\\d\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}, \tag{5.5}$$

which yields three equations

$$\begin{aligned} x+a+b+d &= 0\\ y+a+c+d &= 0\\ z+b+c+d &= 0. \end{aligned}$$

Because we are working in, F, we have $-x = x$ for all x, and the equations becomes

$$\begin{aligned} x &= a+b+d\\ y &= a+c+d\\ z &= b+c+d. \end{aligned} \tag{5.6}$$

The variables *a, b, c, d* may independently take either value, 0 or 1, in F. For this reason they are called *free variables*. There are 2^4 choices for the values of the four free variables. The variables *x, y, z* are determined by these choices and are therefore called *determined variables*. The values of the free variables are called *information bits*, while the values of determined variables are called *check bits*.

We have shown that $|A| = 16$. It remains to prove $d(A) \geq 3$, which we do by showing that A corrects one error. Suppose that a codeword $c \in A$ is sent and one error occurs. Assuming that the error occurs in the ith component, we represent the error by a vector consisting of a single 1 in the ith position and 0′s in all other positions:

$$e = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow i\text{th position}$$

The received vector is $c + e$ (which differs from c in just the ith component), and it is the decoder's job to determine the position in which the error has occurred. This is done by exploiting properties of the matrix H. We multiply H by $c + e$:

$$\begin{aligned} H(c+e) &= Hc + He \\ &= 0 + He \qquad \text{(by definition of } A\text{)} \\ &= He. \end{aligned}$$

Since e has only one nonzero row, the product He consists of the column of H corresponding to this row position. In other words, He equals the ith column of H. Because of the way H is constructed, this is the number i in binary. Thus, when we compute $H(c + e)$ the position of the error is revealed (in binary). We have demonstrated that A corrects one error, so $d(A) \geq 2 \cdot 1 + 1 = 3$.

We have already remarked that A is a perfect code. As a sphere packing, A may be pictured as the centers of the spheres (in Euclidean space) referred to in the introduction to this chapter. As we have already noted, every element of F^3 lies in exactly one sphere. This Hamming code (of dimension 7) has rate $r(A) = (\log_2 16)/7 = 4/7$. In general, the Hamming code of dimension n has rate

$$r(A) = \frac{\log_2 2^m}{n} = \frac{m}{n} = \frac{2^r - 1 - r}{2^r - 1}, \tag{5.7}$$

which tends to 1 as r tends to infinity. Thus, the Hamming codes are a family of 1-error-correcting codes with arbitrarily good rate.

The equations of the previous section allow us to determine the 16 codewords of the (7, 16, 3) Hamming code. They are listed below:

$$\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\end{bmatrix}\quad\begin{bmatrix}1\\1\\0\\1\\0\\0\\1\end{bmatrix}\quad\begin{bmatrix}0\\1\\0\\1\\0\\1\\0\end{bmatrix}\quad\begin{bmatrix}1\\0\\0\\0\\0\\1\\1\end{bmatrix}\quad\begin{bmatrix}1\\0\\0\\1\\1\\0\\0\end{bmatrix}\quad\begin{bmatrix}0\\1\\0\\0\\1\\0\\1\end{bmatrix}\quad\begin{bmatrix}1\\1\\0\\0\\1\\1\\0\end{bmatrix}\quad\begin{bmatrix}0\\0\\0\\1\\1\\1\\1\end{bmatrix}$$

$$\begin{bmatrix}1\\1\\1\\0\\0\\0\\0\end{bmatrix}\quad\begin{bmatrix}0\\0\\1\\1\\0\\0\\1\end{bmatrix}\quad\begin{bmatrix}1\\0\\1\\1\\0\\1\\0\end{bmatrix}\quad\begin{bmatrix}0\\1\\1\\0\\0\\1\\1\end{bmatrix}\quad\begin{bmatrix}0\\1\\1\\1\\1\\0\\0\end{bmatrix}\quad\begin{bmatrix}1\\0\\1\\0\\1\\0\\1\end{bmatrix}\quad\begin{bmatrix}0\\0\\1\\0\\1\\1\\0\end{bmatrix}\quad\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

Since $d(x, y) = w(x - y)$ for any $x, y \in F^n$, the distance of a linear code equals the minimum weight of a nonzero codeword. Table 5.1 tallies the words of the (7, 16, 3) Hamming code according to weight.

Table 5.1 Weight distribution of the Hamming code.

weight	0	3	4	7
number of words	1	7	7	1

The symmetry of the weight distribution is due to the fact that A is a linear code containing the all-1 codeword. Thus, if $v \in A$, then also $v^c \in A$ (where v^c is the binary complement of v) and $w(v^c) = 7 - w(v)$.

The seven Hamming codewords of weight 3 give rise to a finite geometry called the *Fano Configuration* (FC). The Fano Configuration has seven points, 1, 2, 3, 4, 5, 6, 7, corresponding to the seven components of the code vectors of A. Each codeword of weight 3 contains, by definition, three 1′s. The three points corresponding to the 1′s are joined by a line, called an *edge* of FC. The edges are 246, 167, 145, 257, 123, 347, and 356.

EXERCISES

5.10 Explain how sixteen 7-dimensional unit hyperspheres cap be arranged so that each hypersphere is tangent to exactly seven of the other hyperspheres.

5.11 How reliable is the (7, 16, 3) Hamming code? Suppose that $q = 0.05$ (the probability of an error in a single bit of the message). Compare the probability that exactly one error occurs (and hence the code corrects the error) with the probability of an error occurring if no code is used.

5.12 Find a code $A \subseteq \mathrm{GF}(2)^8$ with $d(A) = 4$ and $r(A) = 1/2$.

5.13 Let A be the $(15, 2^{11}, 4)$ Hamming code. Suppose that $v \in A$ is sent, at most one error occurs, and $w = 101000000000000$ is received. Find v.

5.14 Show how to create a cyclic (7, 16, 3) Hamming code by making the transformation [*xyazbcd*] → [*yxzabdc*]. In a cyclic code, cyclic shifts of a codewords are also codewords.

5.15 Use a computer to generate the elements of the $(15, 2^{11}, 4)$ Hamming code.

5.16 Seven players play a game as a team. They are in a room wearing hats of one of two colors, white or black. Each person can see the hats of the others but not his or her own hat. Each person's hat color is chosen randomly, with probability of white and black hats equally likely and independent of the other players' hats. Simultaneously, the players guess their hat colors or "pass" if they don't want to guess. The team wins if at least one person guesses correctly and no person guesses incorrectly. The players are allowed to discuss possible strategies before they enter the room. What is their best strategy and what is their probability of winning?

The general version of this problem (with n players) is due to Todd Ebert, who proposed it in his Ph.D. thesis at the University of California at Santa Barbara in 1998.

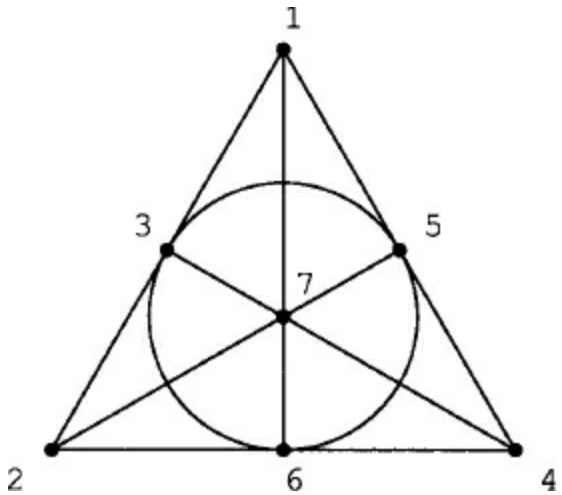
5.17 Prove that the probability that the determinant of a random $n \times n$ matrix over a finite field is 0 tends to a constant as n tends to infinity.

5.4 The Fano Configuration

The simplest combinatorial construction is known as the Fano Configuration, named after the geometer Gino Fano (1871–1952)[1]. The Fano Configuration is the prototypical example of many types of structures, such as projective planes, block designs, and difference sets. Let's look at the configuration and observe some of its properties.

The Fano Configuration (FC) is shown in Figure 5.2.

Figure 5.2 The Fano Configuration (FC).



The points of FC are labeled 1, 2, 3, 4, 5, 6, and 7. The lines are just unordered triples of points (e.g., {1, 2, 3}); they have no Euclidean meaning Hence, the points may be located anywhere in space, and the lines may be drawn straight or curved and may cross arbitrarily.

We observe that FC has the following properties:

1. There are seven points.

2. There are seven lines.

3. Every line contains three points.

4. Every point lies on three lines.

5. Every two points lie on exactly one line.

6. Every two lines intersect in exactly one point.

The above properties occur in pairs called *duals*. If the words "point" and "line" are interchanged, each property is transformed into its dual property.

FC has many fascinating properties. For instance, with its edges properly oriented, FC represents the multiplication table for the Cayley algebra. Here is a simple application.

EXAMPLE 5.2

Seven students are asked to evaluate a set of seven textbooks, making comparisons between pairs of textbooks. While it is possible for every student to read every book and write comparisons between each pair, this would be time consuming. Instead, each student receives three books to read, according to the FC diagram. Number the students 1, 2, 3, 4, 5, 6, 7. The books correspond to the lines and therefore can be numbered {1, 2, 3}, {2, 4, 6}, etc. Then, for example, student 2 receives books {1, 2, 3}, {2, 4, 6}, and {2, 5, 7}. Each student writes a comparison between the three pairs of books that he or she reads. Thus, each student must write only three comparisons, instead of 21.

We next determine the automorphism group of FC, denoted Aut FC. This automorphism group is the group of permutations of the vertices of FC that preserve collinearity. For example, if we rotate FC clockwise by one-third of a circle, collinearity is preserved. This permutation is written in cycle notation as (142)(356)(7).

We begin by calculating the order of Aut FC. It is evident from Figure 5.2 that all seven vertices are equivalent in terms of collinearity. Therefore, vertex 1 may be sent by an automorphism to any of the seven vertices. Suppose that 1 is mapped to 1′. Vertex 2 may be mapped to any of the remaining six vertices. Suppose that 2 is mapped to 2′. In order to preserve collinearity, vertex 3 must be mapped to the unique point collinear with 1′ and 2′. Call this point 3′. Vertex 4 is not on line 123, so its image 4′ can be any of the remaining four vertices. Finally, the images of the other points are all determined by collinearity: 5 is collinear with 1 and 4; 6 is collinear with 2 and 4; and 7 is collinear with 3 and 4. Hence, there are $7 \cdot 6 \cdot 4 = 168$ automorphisms.

Now we know that Aut FC is a group of order 168, but what group? Is it abelian? Cyclic? Simple? We will show that Aut FC is isomorphic to the group of invertible 3×3 matrices over F.

Here is some background about groups of matrices. The *general linear group* GL(n, q) is the set of invertible $n \times n$ matrices with coefficients in the Galois field GF(q) of order $q = p^k$, where p is a prime, under matrix multiplication. The order of GL(n, q) is readily determined. There are $q^n - 1$ choices for the first row of an invertible $n \times n$ matrix (the all-0 row is excluded). Having chosen the first row, the second row may be any of the q^n possible n-tuples except the q scalar multiples of the first row. Hence, there are $q^n - q$ choices for the second row. Similarly, the third row may be any of the q^n n-tuples except linear combinations of the first two rows. There are $q^n - q^2$ choices. Continuing in this manner, we arrive at the total number of invertible matrices:

$$|\mathrm{GL}(n, q)| = (q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}). \tag{5.8}$$

For example, GL(2, 2) has six elements and is in fact isomorphic to S_3. The group Aut FC is

isomorphic to GL(3, 2).

The *special linear group* SL(n, q) is a subgroup of GL(n, q) consisting of those $n \times n$ invertible matrices with entries from GF(q) and determinant 1. We claim that SL(n, q) is a normal subgroup of GL(n, q). For if $M \in$ SL(n, q) and $N \in$ GL(n, q), then

$$\det(NMN^{-1}) = \det N \det M \det N^{-1} = \det N \det N^{-1} = 1.$$

Now consider the homomorphism

$$\begin{aligned} f\colon \mathrm{GL}(n,q) &\longrightarrow \mathrm{GF}(q)\setminus\{0\} \\ M &\longmapsto \det M. \end{aligned}$$

The kernel of f is SL(n, q) and the homomorphism is clearly onto. Therefore, by the first homomorphism theorem for groups,

$$\mathrm{GL}(n,q)/\mathrm{SL}(n,q) \simeq \mathrm{GF}(q)\setminus\{0\}, \tag{5.9}$$

and hence,

$$|\mathrm{SL}(n,q)| = \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{n-1})}{q-1}. \tag{5.10}$$

Projective versions of the groups GL(n, q) and SL(n, q) are obtained as the quotient groups of these groups by their centers. The *projective general linear group* PGL(n, q) is GL(n, q)/Z(GL(n, q)) and the *projective special linear group* PSL(n, q) is SL(n, q)/Z(SL(n, q)). In the exercises, the reader is asked to prove the following formulas:

$$|\mathrm{PGL}(n,q)| = \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{n-1})}{q-1} \tag{5.11}$$

and

$$|\mathrm{PSL}(n,q)| = \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{n-1})}{d(q-1)}, \tag{5.12}$$

where $d = \gcd(n, q-1)$.

For all $n \geq 1$, we have

$$\mathrm{GL}(n,2) \simeq \mathrm{SL}(n,2) \simeq \mathrm{PGL}(n,2) \simeq \mathrm{PSL}(n,2).$$

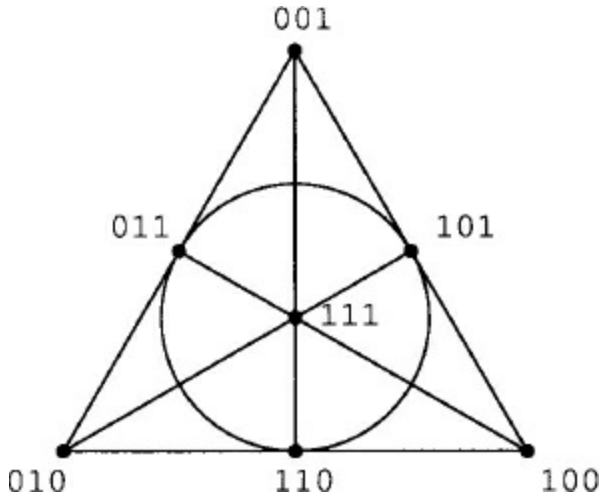
When $n = 3$ we obtain

$$|\mathrm{GL}(3,2)| = \frac{(2^3-1)(2^3-2)(2^3-2^2)}{2-1} = 7\cdot 6\cdot 4 = 168.$$

We will now show that Aut FC is isomorphic to GL(3, 2).

We label the vertices of FC with the seven nonzero vectors in F^3, as in Figure 5.3. This labeling is derived from the labeling of Figure 5.2 by assigning to each vertex i the vector that represents the number i in binary. The vectors have been chosen so that v_1, v_2, v_3 are collinear if and only if $v_1 + v_2 + v_3 = 0$ (in F^3). The matrix group GL(3, 2) acts on the vectors of FC in the obvious way: $v \mapsto vM$. It is easy to check that this action preserves collinearity: $v_1 + v_2 + v_3 = 0$ if and only if $(v_1 + v_2 + v_3)M = 0 \cdot M$, which is true if and only if $v_1M + v_2M + v_3M = 0$. Therefore, Aut FC is isomorphic to the group of 3×3 invertible matrices over F.

Figure 5.3 A vector representation of FC.



It is well known (see [23]) that PSL(n, q) is a simple group (a group with no nontrivial normal subgroups) for all $n \geq 2$ except $n = 2$ and $q = 2$ or $q = 3$. Because there is only one nonabelian group of order 6, it follows that PSL(2, 2) $\simeq S_3$. It is also easy to show that PSL(2, 3) $\simeq A_4$. The simple groups are crucial to the study of algebra. Some of them are difficult to describe, although we have shown that PSL(3, 2) ($\simeq$ GL(3, 2)) has a nice geometric model.

Let us find generators for Aut FC. By direct calculation, we see that the rows of M are the images of the binary representations of 1, 2, and 4. Let

$$S = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

We observe from Figure 5.3 that T yields a reflection of FC around the 374 axis while S yields the 7-cycle (1 5 6 7 2 4 3). We will show that all 168 matrices in the automorphism group of FC are generated by combinations of S and T.

The "visible automorphisms" of FC are the symmetries of its equilateral triangular shape given by the elements of the symmetric group S_3. These symmetries are combinations of S and T, for T is a reflection of the triangle (i.e., a transposition of two of its vertices) and $(S^2TS^2)^2T$ is a rotation of the triangle by one-third of a circle; all symmetries of the triangle are combinations of a reflection and a rotation. Furthermore, STS^2T is a transvection (the identity matrix with an extra 1 in an off-diagonal position). From the transvection, we can produce all transvections via conjugation by permutations. Using elementary row operations, all invertible matrices can be formed from permutation matrices and transvections. Therefore, all invertible matrices are combinations of S and T.

A presentation of the group is

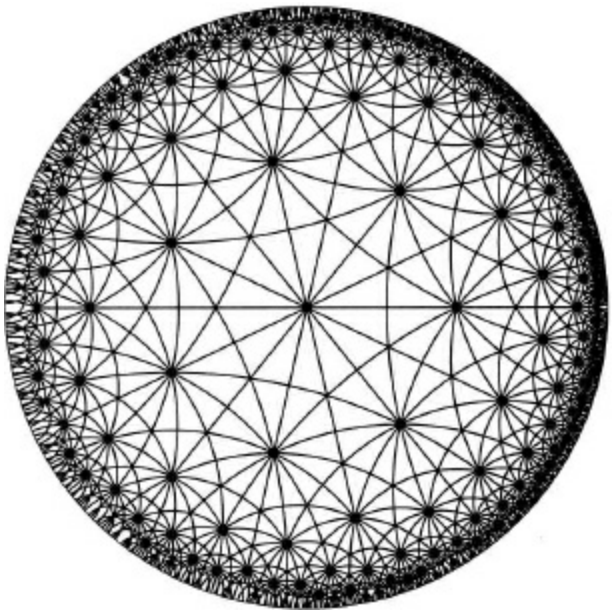
$$G = \langle s, t : s^7 = t^2 = (st)^3 = (s^4t)^4 = 1 \rangle.$$

The matrix ST has order 3 (but isn't a rotation). Let $U = (ST)^{-1}$. Then the group is a homomorphic image of the "triangle group" generated by S, T, and U:

$$\langle s, t, u : s^7 = t^2 = u^3 = stu = 1 \rangle.$$

This is the (infinite) group of symmetries of a tiling of the hyperbolic plane with triangles whose angles are $\pi/2$, $\pi/3$, and $\pi/7$. This tiling is shown in Figure 5.4.

Figure 5.4 A triangle tiling of the hyperbolic plane.



EXERCISES

5.18 Find a design, other than FC, with seven points and seven lines, each line containing three points, and each point on three lines.

5.19 (Sylvester's problem) Can you draw FC in the plane with straight lines? This exercise will show that it is not possible.

Suppose that X is a finite set of points in the plane with the property that every line determined by two points of X contains a third point of X.

(a) Prove that X is collinear.

(b) Show that the assertion in part (a) is false if X is infinite.

5.20 Prove that

$$\frac{(2^n-1)(2^n-2)(2^n-2^2)(2^n-2^3)\cdots(2^n-2^{n-1})}{n!}$$

is an integer for all $n \leq 1$.

5.21 Prove formulas (5.11) and (5.12).

5.22 Show that $(S^2\,TS^2)^2T$ is a rotation of FC by one-third of a circle and STS^2T is a transvection.

5.23 Prove that Gauss's q-binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^n-q)(q^n-q^2)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)(q^k-q^2)\cdots(q^k-q^{k-1})}$$

is equal to the number of k-dimensional subspaces of a vector space of dimension n over a field of size q.

5.24 Show that the elements of GL(3, 2) have orders 1, 2, 3, 4, or 7. How many elements of each order are there?

5.25 What group is Aut($\mathbf{Z}_2 + \mathbf{Z}_2 + \mathbf{Z}_2$)?

5.26 Prove that PSL (7, 2) $\simeq$ GL(3, 2).

Hint: Use a group presentation.

5.27 Prove that the automorphism group of the Hamming code of length $n = 2^r - 1$ is isomorphic

to GL(r, 2).

5.28 Prove that the group

$$G = \langle s, t : s^7 = t^2 = (st)^3 = (s^4t)^4 = 1 \rangle$$

has 168 elements.

Notes

Richard Hamming (1915–1998) developed the Hamming codes in 1947 at Bell Telephone Laboratories in order to solve the problem of glitches in running computer programs. Marcel Golay (1902–1989) did much of the same work independently at the Signal Corps Engineering Laboratories. For a history of their achievements the reader is referred to [27]. Today, error-correcting codes are used in the design of compact discs (CDs), satellite communications, ISBN numbers, and bar code scanners. Error-correcting codes are an important part of the science of information theory introduced by Claude Shannon (1916–2001) in 1941 (also at Bell Labs). The main ideas of information theory are that an information source contains a certain amount of uncertainty (entropy) and that entropy determines the accuracy and amount of information that can be sent over a communication channel. In a memoryless source of English characters, for example, the entropy is thought to be about 4.03 bits (each revealed character conveys about 4.03 bits of information). As a baseline comparison, the entropy of a memoryless source of 27 equally likely symbols (26 letters and a space) is approximately 4.76 bits. (By the way, in a popular word-making board game the distribution of letter tiles gives an entropy value of about 4.32 bits.) However, if knowledge of the English language is used to predict the likelihood of future characters (the source has a memory), then the entropy is believed to be between 0.6 and 1.3 bits. Knowledge of the information structure of a language is useful in the design of machines that translate the spoken word into written text (continuous speech recognition).

A linear code has a *generating matrix*. The (7, 16, 3) Hamming code, as constructed in this chapter, is generated by the 7 × 4 matrix

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The code is the range of the corresponding linear transformation, the set of vectors $Gv \in F^7$, where $v \in F^4$.

For a detailed introduction to error-correcting codes, a good source is [22].

[1] In Fano's geometry, what we call the Fano Configuration was specifically excluded.

CHAPTER 6

COMBINATORIAL DESIGNS

The Fano Configuration FC is the simplest nontrivial example of many types of combinatorial configurations, including t-designs, Steiner systems, block designs, and projective planes. We next explore these designs and investigate their interconnections, paving the way for the construction of the Golay code G_{23}, the only perfect binary code capable of correcting more than one error, and Leech's remarkable 24-dimensional lattice.

6.1 *t*-designs

A t-(v, k, λ) *design* (or *t-design*) consists of a v-set S and a collection C of k-subsets of S, with the property that every t-subset of S is contained in exactly λ members of C. The elements of S are called *points* and the elements of C are called *blocks*.

A t-design is *nontrivial* if $0 < t < k < v$ and not every k-subset of S is a block.

EXAMPLE 6.1

The lines 246, 167, 145, 257, 123, 347, and 356 of FC are the blocks of a 2-(7, 3, 1) design with $S = \{1,\ldots,7\}$.

EXAMPLE 6.2

The complement FC' of FC has the same set of vertices as FC. Its lines are the complements of the lines of FC: 1357, 2345, 2367, 1346, 4567, 1256, 1247. It is easy to verify that the lines of FC' constitute the blocks of a 2-(7, 4, 2) design.

EXAMPLE 6.3

The following sets are the blocks of a 3-(8, 4, 1) design:

1357 2345 2367 1346 4567 1256 1247
2468 1678 1458 2578 1238 3478 3568

These blocks are of two types: (1) the lines of FC' and (2) the lines of FC joined to a new element 8. The *derived* design obtained by removing any point and all the sets not incident with it is equivalent to FC. Conversely, we call the 3−(8, 4, 1) design an *extension* of FC.

EXAMPLE 6.4

A graph with p vertices and q edges is a 0-(p, 2, q) design. An r-regular graph is a 1-(p, 2, r) design. An r-regular k-uniform hypergraph is a 1-(p, k, r) design.

We say that two t-designs are *equivalent* if they can be made the same by relabeling their underlying sets. Each of the nontrivial designs above is unique up to this equivalence. One reason for the uniqueness is that many parameters of a design are determined by the following theorem.

Parameter theorem. Given a t-(v, k, λ) design with point set S and block set C and $0 \le i \le t$, there exists a constant λ_i such that every i-set of S lies in exactly λ_i elements of C. Therefore, the design is an i-(v, k, λ_i) design. Furthermore, λ_i satisfies $\lambda_i \binom{k-i}{t-i} = \lambda \binom{v-i}{t-i}$.

Proof. Let X be a fixed subset of S with $|X| = i$, and consider the ordered pairs (T, K) with $|T| = t$, $K \in C$ and $X \subseteq T \subseteq K$. As in the proof of Burnside's lemma, we count the ordered pairs in two ways (from the perspective of each coordinate) to obtain $\lambda_i \binom{k-i}{t-i} = \lambda \binom{v-i}{t-i}$, a relation independent of X.

We solve the parameter equation for λ_i:

$$\lambda_i = \lambda \frac{\binom{v-i}{t-i}}{\binom{k-i}{t-i}}. \tag{6.1}$$

Putting in the value $i = 0$, we obtain the number b of blocks in C:

$$b = \lambda_0 = \lambda \frac{\binom{v}{t}}{\binom{k}{t}}. \tag{6.2}$$

Putting in the value $i = 1$, we obtain the number r of times each element of S occurs in a block:

$$r = \lambda_1 = \lambda \frac{\binom{v-1}{t-1}}{\binom{k-1}{t-1}}. \tag{6.3}$$

The reader should verify the formulas for b and r in the above examples.

An $S(t, k, v)$ *Steiner system* is a t-design with $\lambda = 1$. Later, we construct the Golay code G_{23} via a related code that contains an $S(5, 8, 24)$ Steiner system. No Steiner system is known with $t > 5$. The only known Steiner systems with $t = 5$ are $S(5, 6, 12)$, $S(5, 6, 24)$, $(5, 6, 36)$, $S(5, 6, 48)$, $S(5, 6, 72)$, $S(5, 6, 84)$, $S(5, 6, 108)$, $S(5, 6, 132)$, $S(5, 6, 168)$, $S(5, 6, 244)$, $S(5, 7, 28)$, and $S(5, 8, 24)$ designs, and the only known Steiner systems with $t = 4$ are derived from these designs.

Open problem. Determine whether there is a Steiner system with $t > 5$.

A *Steiner triple system* is a Steiner system with $k = 3$.

EXAMPLE 6.5

FC is a $S(2, 3, 7)$ Steiner triple system.

EXAMPLE 6.6

An $S(2, 3, 9)$ Steiner triple system is given by the set of blocks

$$\{123, 456, 789, 147, 168, 159, 258, 369, 249, 357, 267, 348\}.$$

This Steiner system is equivalent to the set of nonideal points and lines of the projective plane of order 3.

EXAMPLE 6.7

Show that the weight 3 codewords of the Hamming code of length 15 form a Steiner triple system. What are its parameters?

EXAMPLE 6.8

The 81 cards of the SET® game are the points of an $S(2, 3, 81)$ Steiner triple system. The blocks are the 1080 possible sets. Each block contains three points and each point is contained in exactly 40 blocks.

We can define further parameters for a t-(v, k, λ) design.

Double-parameter theorem. Given a t-(v, k, λ) design with points S and blocks C, and i and j nonnegative integers satisfying $i + j \leq t$, there exists a constant λ_{ij} such that the number of blocks which contain all the elements of any fixed i-set of S and omit all the elements of any fixed j-set of S (the i-set having no elements in common with the j-set) is exactly λ_{ij}. Furthermore, λ_{ij} satisfies

$$\lambda_{ij}\binom{v-t}{k-t} = \lambda\binom{v-i-j}{k-i}.$$

Proof. From the parameter theorem and the inclusion–exclusion principle, we obtain

$$\lambda_{ij} = \sum_{s=0}^{j} \lambda(-1)^s \frac{\binom{v-(i+s)}{k-(i+s)}}{\binom{v-t}{k-t}}\binom{j}{s}.$$

It follows that

$$\begin{aligned}
\lambda_{ij} &= \frac{\lambda}{\binom{v-t}{k-t}} \sum_{s=0}^{j} (-1)^s \binom{v-i-s}{k-i-s}\binom{j}{s} \\
&= \frac{\lambda}{\binom{v-t}{k-t}} (-1)^{k-i} \sum_{s=0}^{j} \binom{-v+k-1}{k-i-s}\binom{j}{s} \\
&= \frac{\lambda}{\binom{v-t}{k-t}} (-1)^{k-i} \binom{-v+k-1+j}{-v+k-1-i+v+1} \\
&= \frac{\lambda}{\binom{v-t}{k-t}} (-1)^{k-i} \binom{-v+k-1+j}{k-i} \\
&= \frac{\lambda}{\binom{v-t}{k-t}} \binom{v-i-j}{k-i}.
\end{aligned}$$

EXAMPLE 6.9

Recall that FC is a 2-(7, 3, 1) design. Applying the double-parameter theorem, we determine the constants $\lambda_{00} = 7$, $\lambda_{10} = 3$, $\lambda_{01} = 4$, $\lambda_{20} = 1$, $\lambda_{11} = 2$, $\lambda_{02} = 2$. For example, the relation $\lambda_{11} = 2$ says that, given any points x and y of FC, there are precisely two lines which contain x and omit y.

EXERCISES

6.1 Draw the $S(2, 3, 9)$ Steiner system.

6.2 Construct a $2-(21, 5, 1)$ design.

6.3 Prove that in an $S(4, 5, 11)$ Steiner system no two blocks are disjoint. Prove that in an $S(5, 6, 12)$ Steiner system the complement of any block is a block.

Hint: Let $\{a, b, c, d, e\}$ be a block in the $S(4, 5, 11)$ Steiner system. Let A be the set of blocks containing a, B the set of blocks containing b, etc. Use the inclusion–exclusion principle and knowledge of the values of λ_i to find $|A \cup B \cup C \cup D \cup E|$. Solve the problem about $S(5, 6, 12)$ similarly.

6.4 Prove that the double parameters λ_{ij} satisfy the relations $\lambda_{i0} = \lambda_i$ and $\lambda_{(i-1)j} = \lambda_{(i-1)(j-1)} - \lambda_{i(j-1)}$. Show that from these relations the values of λ_{ij} for all $i+j \leq t$ can be calculated.

6.5 Show that a $2-(2n + 1, n, \lambda)$ design can be extended (by adding one element) to a $3-(2n + 2, n + 1, \lambda)$ design.

6.2 Block designs

A *balanced incomplete block design* (BIBD) is a nontrivial $2-(v, k, \lambda)$ design. The parameter theorem shows that there is a number r such that each element of the set S occurs in exactly r blocks. Letting $v = |S|$, we rephrase the definition of balanced incomplete block design. A (v, b, r, k, λ) BIBD is a family C of b subsets (*blocks* or *lines* or *edges*) of a set S of v elements (*points* or *vertices*) such that:

1. Each point of S lies in exactly r blocks.

2. Each block has k points.

3. Each pair of points of S occur together in λ blocks (the "balance" condition).

4. Not every k-set of S is a block (the "incompleteness" condition).

The nontriviality condition becomes $2 < k < v$.

EXAMPLE 6.10

FC is a $(7, 7, 3, 3, 1)$ BIBD. FC′ is a $(7, 7, 4, 4, 2)$ BIBD. The $S(2, 3, 9)$ Steiner system is a $(9, 12, 4, 3, 1)$ BIBD.

Parameter theorem for block designs. In a (v, b, r, k, λ) BIBD, $bk = vr$ and $r(k - 1) = \lambda(v - 1)$.

Proof. These relations follow from the parameter theorem for t-designs upon letting $t = 2$. The first result is obtained by dividing the relation (6.2) by the relation (6.3). The second result is immediate from (6.3).

The relations of the theorem can be proved without using the parameter theorem. To prove the first relation, note that the total number of incidences between vertices and blocks is bk (from the point of view of the blocks) and also vr (from the point of view of the vertices). Therefore $bk = vr$. To prove

the second relation, note that the number of times a particular element occurs in pairs with other elements is $r(k-1)$. But it is also $\lambda(v-1)$, the number of elements with which the particular element may be paired multiplied by the number of times each pair occurs. Therefore $r(k-1) = \lambda(v-1)$.

The incidence relation between blocks and vertices can be displayed in an *incidence matrix* $A = [a_{ij}]_{b \times v}$. We choose orderings of the points and the blocks and let $a_{ij} = 1$ if the jth point is an element of the ith block and $a_{ij} = 0$ otherwise. For instance, the points 1,…,7 and the blocks 246, 167, 145, 257, 123, 347, and 356 of FC are represented by the incidence matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

Theorem. For any (v, b, r, k, λ) BIBD:

(6.4) $\det(A^t A) = rk(r-\lambda)^{v-1}$.

Proof. From the block size and balance condition of a BIBD, it follows that

$$\det(A^t A) = \begin{vmatrix} r & \lambda & \dots & \lambda \\ \lambda & r & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda \\ \lambda & \dots & \lambda & r \end{vmatrix}. \tag{6.5}$$

Subtracting the first row from the others, then replacing the first column with the sum of all the columns, we obtain

$$\begin{aligned} \det(A^t A) &= \begin{vmatrix} r+\lambda(v-1) & \lambda & \lambda & \dots & \lambda \\ 0 & r-\lambda & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & 0 & r-\lambda & 0 \\ 0 & \cdots & 0 & 0 & r-\lambda \end{vmatrix} \\ &= [r+(v-1)\lambda](r-\lambda)^{v-1} \\ &= rk(r-\lambda)^{v-1}. \end{aligned}$$

Fisher's inequality. In any BIBD, $b \geq v$.

Proof. Let A be an incidence matrix of the (v, b, r, k, λ) BIBD. Thus, A is a matrix of dimensions $b \times v$. Let I be the $v \times v$ identity matrix and J the $v \times v$ all-1′s matrix. From the previous theorem it follows that $A^tA = \lambda J + (r-\lambda)I$ and $\det A^tA = rk(r-\lambda)^{v-1}$. The relation $r(k-1) = \lambda(v-1)$ implies that $r > \lambda$, which means that $\det(A^tA) \neq 0$. Hence, $v = \text{rank}(A^tA) \leq \min\{v, b\} \leq b$.

The extreme case $v = b$ gives rise to an interesting subclass of block designs. A (v, k, λ) *square*

block design (SBD) is a (v, b, r, k, λ) BIBD in which $v = b$ (and hence also $k = r$). Square block designs are usually called "symmetric block designs," but this is probably not the best term for them as their incidence matrices are not symmetric. We believe that the term "square block design" is a better choice. Note that in a (v, k, λ) SBD we have $k(k-1) = \lambda(v-1)$.

EXAMPLE 6.11

FC is a (7, 3, 1) SBD and FC' is a (7, 4, 2) SBD.

Theorem. The incidence matrix A of a (v, k, λ) SBD is normal: $A^tA = AA^t$. Thus, any two distinct blocks intersect in exactly λ elements.

Proof. We have $A^tA = (k-\lambda)I + \lambda J$. Because $\det(A^tA) \neq 0$, it follows that $\det A \neq 0$, so A^{-1} exists. Therefore

$$AA^t = AA^tAA^{-1} = A(k-\lambda)IA^{-1} + A\lambda JA^{-1} = (k-\lambda)I + \lambda AJA^{-1}.$$

Also, $JA = kJ$ implies that $JA^{-1} = k^{-1}J$. Hence $AA^t = (k-\lambda)I + \lambda J$. An interpretation of this matrix product yields the desired intersection property.

If A is the incidence matrix of an SBD, then $(\det A)^2 = k^2(k-\lambda)^{v-1}$, and so $\det A = k(k-\lambda)^{(v-1)/2}$. It follows that if v is even, then $k - \lambda$ is a perfect square. This is the first part of the Bruck–Chowla–Ryser theorem, stated below. For a proof of the second part, see [12].

Bruck–Chowla–Ryser theorem (1949). Suppose that a (v, k, λ) SBD exists. Then the following two statements hold:

1. If v is even, then $k - \lambda$ is a perfect square.

2. If v is odd, then the equation

$$x^2 = (k-\lambda)y^2 + (-1)^{(v-1)/2}\lambda z^2$$

has a solution in integers x, y, and z, not all zero.

EXAMPLE 6.12

There is no (22, 7, 2) SBD, since $k - \lambda = 5$, which is not a perfect square.

Let D be a BIBD. The *complement* D' of D is obtained by switching 0 and 1 in an incidence matrix of D. It is easy to check that the complement of a (v, b, r, k, λ) BIBD is a $(v, b, b-r, v-k, b-2r+\lambda)$ BIBD. Specifically, the complement of a (v, k, λ) SBD is a $(v, v-k, v-2k+\lambda)$ SBD.

EXAMPLE 6.13

The $S(2, 3, 9)$ Steiner system is a (9, 12, 4, 3, 1) BIBD whose complement is a (9, 12, 8, 6, 5) BIBD.

If $p \equiv 3 \pmod 4$, then we can construct a $\left(p, \frac{p+1}{2}, \frac{p+1}{4}\right)$ square design via the set R_p of quadratic residues modulo p. If p is any prime greater than 2, then the map $f: \mathbf{Z}_p^* \to R_p$ with $f(x) = x^2$ is an epimorphism with kernel $\{-1, 1\}$, from which it follows by the first homomorphism theorem for groups that $|R_p| = (p-1)/2$. Let N_p be the set of quadratic nonresidues modulo p, so that $|N_p| = (p -$

1)/2. The *Legendre symbol* (x / p) is defined as

$$\left(\frac{x}{p}\right) = \begin{cases} 0 & \text{if } x \equiv 0 \pmod{p} \\ 1 & \text{if } x \in R_p \\ -1 & \text{if } x \in N_p. \end{cases} \tag{6.6}$$

Because $|R_p| = |N_p|$, we have $\Sigma(x/p) = 0$ for any sum over a complete residue system modulo p.

Assuming that $p \equiv 3 \pmod 4$, let A be the $p \times p$ circulant binary matrix whose first row is the characteristic vector of $R_p \cup \{0\}$ and whose other rows are successive one unit shifts to the right of the first row. We claim that A is the incidence matrix of a $(p, \frac{p+1}{2}, \frac{p+1}{4})$ SBD. Evidently, $v = b = p$ and $k = r = \frac{p+1}{2}$. We only need to check that the dot product of any two distinct rows is $\frac{p+1}{4}$. The dot product of two rows which differ by a shift of j units to the right is

$$\lambda = \frac{1}{2}\left[\left(\frac{j}{p}\right)+1\right] + \frac{1}{2}\left[\left(\frac{-j}{p}\right)+1\right] + \frac{1}{4}\sum_{x\in S}\left[\left(\frac{x}{p}\right)+1\right]\left[\left(\frac{x+j}{p}\right)+1\right],$$

where S is a complete residue system modulo p except for the values 0 and $-j$. Since p is congruent to 3 modulo 4, it follows that -1 is a quadratic nonresidue modulo p. Therefore, exactly one of j and $-j$ is a quadratic residue and the other is a quadratic nonresidue. Letting x^{-1} be the multiplicative inverse of x, we have

$$\begin{aligned} \lambda &= 1 + \frac{1}{4}\sum_{x\in S}\left[\left(\frac{x}{p}\right)\left(\frac{x+j}{p}\right) + \left(\frac{x+j}{p}\right) + \left(\frac{x}{p}\right) + 1\right] \\ &= 1 + \frac{1}{4}\sum_{x\in S}\left(\frac{x^2}{p}\right)\left(\frac{1+jx^{-1}}{p}\right) + \frac{1}{4}\sum_{x\in S}\left(\frac{x+j}{p}\right) + \frac{1}{4}\sum_{x\in S}\left(\frac{x}{p}\right) + \frac{1}{4}\sum_{x\in S}1 \\ &= 1 + \frac{1}{4}\sum_{x\in S}\left(\frac{1+jx^{-1}}{p}\right) - \frac{1}{4}\left(\frac{j}{p}\right) - \frac{1}{4}\left(\frac{-j}{p}\right) + \frac{1}{4}(p-2) \\ &= \frac{p+2}{4} + \frac{1}{4}\sum_{x\in S}\left(\frac{1+jx^{-1}}{p}\right). \end{aligned}$$

Because x^{-1} takes all values except 0 and $-j^{-1}$, it follows that $1 + jx^{-1}$ takes all values except 1 and 0. Therefore

$$\sum_{x\in S}\left(\frac{1+jx^{-1}}{p}\right) = -1$$

and

$$\lambda = \frac{p+2}{4} - \frac{1}{4} = \frac{p+1}{4}.$$

For example, with $p = 7$, the construction furnishes a (7, 4, 2) design equivalent to FC′.

This construction produces square designs with large values of λ. Such designs are equivalent to Hadamard designs. At the opposite extreme, the next section deals with $\lambda = 1$ designs, which are called projective planes.

EXERCISES

6.6 Prove that the complement of a (v, b, r, k, λ) BIBD is a $(v, b, b - r, v - k, b - 2r + \lambda)$ BIBD.

6.7 Find a circulant incidence matrix for FC.

6.8 Construct an (11, 6, 3) SBD.

6.9 Construct a (37, 9, 2) SBD

Hint: Let one set be the nonzero fourth powers modulo 37.

6.10 Construct a (16, 6, 2) SBD.

Hint: Let one set be the cubes in a field of 16 elements.

6.11 Prove the nonuniform Fisher inequality (R. C. Bose, 1949): Let $C_1,\ldots,C_m \subseteq \{1,\ldots,n\}$ and suppose that $|C_i \cap C_j| = \lambda$, where $1 \leq \lambda < n$. Then $m \leq n$.

6.12 (E. R. Berlekamp, 1969) Prove that if $C_1,\ldots,C_n$ are subsets of a t-set such that $|C_i|$ is odd for all i and $|C_i \cap C_j|$ is even for all $i \neq j$, then $n \leq t$.

Hint: Show that the characteristic vectors of the C_i are, linearly independent over the field {0, 1}.

6.13 Use the previous two exercises to prove the following constructive lower bound for diagonal Ramsey numbers (Z. Nagy, 1972): $R(t + 1, t + 1) > \binom{t}{3}$.

Hint: Let the vertices of K_v, where $v = \binom{t}{3}$, be the 3-subsets of $\{1,\ldots,t\}$. Color the edge $\{X, Y\}$ red $|X \cap Y| = 1$ and green if $|X \cap Y| = 0$ or 2.

6.3 Projective planes

In a (v, k, λ) SBD, suppose that $\lambda = 1$, and set $n = k - 1$. Then from the relation $k(k - 1) = \lambda(v - 1)$, we have $v = n^2 + n + 1$. Such a design is called a *finite projective plane* π_n of order n. Thus, a π_n is an $(n^2 + n + 1, n + 1, 1)$ SBD. The elements of π_n are called *points* and the blocks are called *lines*. As a square design, a finite projective plane of order n has the following properties:

1. There are $n^2 + n + 1$ points.

2. There are $n^2 + n + 1$ lines.

3. Every line is incident with $n + 1$ points.

4. Every point is incident with is $n + 1$ lines.

5. Every two points determine a unique line.

6. Each pair of lines determines a unique point.

As we have seen with FC, these properties occur in pairs called *duals* properties. If the words "point" and "line" are interchanged, each property is transformed into its dual property.

EXAMPLE 6.14 A projective plane of order 2

FC is a projective plane of order 2.

Another description of projective planes comes from linear algebra. For FC, let $F = \{0, 1\}$ be the two-element field, and let $V = F^3 = \{(x, y, z) : x, y, z \in F\}$, the three-dimensional vector space over F. The points of the projective plane are the seven 1-dimensional subspaces of V, and the lines are the seven 2-dimensional subspaces of V. The reader can check that the above six properties hold.

Theorem. A projective plane π_n exists for every prime power n.

Proof. Let $F = \mathrm{GF}(n)$, the Galois field of order n. Let the set of points of the plane be

$$S = \{(i,j): i,j \in F\} \cup \{i: i \in F\} \cup \{\infty\}.$$

The points $i \in F$ are called *ideal points*. The point ∞ is called the *point at infinity*. The lines are

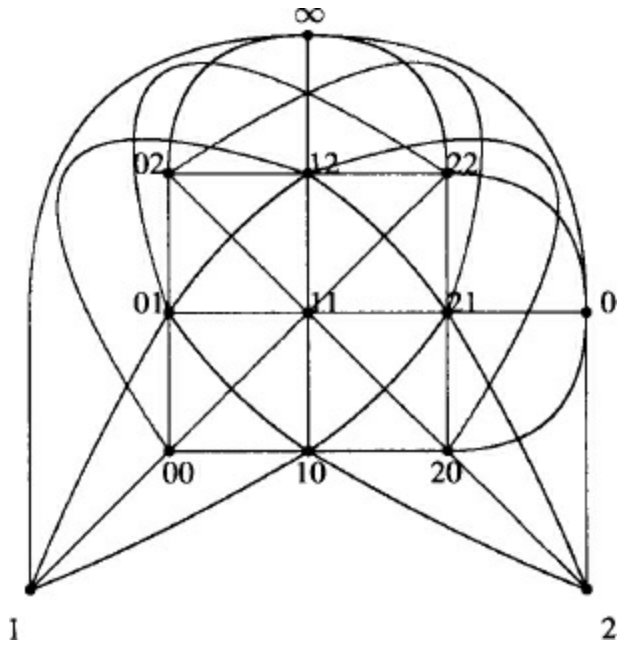
$$l_{m,b} = \{(x,y) \in F^2: y = mx + b\} \cup \{m\}, \quad m, b \in F$$
$$l_k = \{(k,y): y \in F\} \cup \{\infty\}, \quad k \in F$$
$$l_\infty = \{m: m \in F\} \cup \{\infty\}.$$

The line l_∞ is called the *line at infinity*. These $n^2 + n + 1$ points and $n^2 + n + 1$ lines constitute a π_n. We need only check that the conditions for a $(n^2 + n + 1, n^2 + n + 1, n + 1, n + 1, 1)$ BIBD are satisfied. Each point (i, j) lies on exactly $n + 1$ lines, namely, the lines $l_{m,j - mi}$, where $m \in F$, and l_i. The reader should check that the ideal points and the point at infinity each lie on $n + 1$ lines. That each line contains $n + 1$ points can be seen from the definitions. We leave it to the reader to check that each pair of points determines exactly one line.

EXAMPLE 6.15 A projective plane of order 3

Let us use the description in the above proof to construct a projective plane of order 3 (Figure 6.1). The appropriate base field is $F = \mathrm{GF}(3)$, whose elements are 0, 1, 2. The 13 points are (i, j), with $0 \le i, j \le 2$, which constitute the square array of the figure, and the ideal points i, with $0 \le i \le 2$, and ∞, which are placed to the side. The lines are $l_{m,b}$, where $0 \le m, b \le 2$; l_k, where $0 \le k \le 2$; and l_∞.

Figure 6.1 A projective plane of order 3.



By the Bruck–Chowla–Ryser theorem, if there is a projective plane of order n and if $n \equiv 1, 2 \pmod 4$, then there is a solution in integers to the equation

$$x^2 = ny^2 - z^2.$$

It follows that

$$(x/y)^2 + (z/y)^2 = n,$$

i.e., n is expressible as the sum of the squares of two rational numbers. From elementary number theory, it follows that n is the sum of two squares of integers. Hence, there is no projective plane of order 6 or 14. However, $10 = 3^2 + 1^2$, so the Bruck–Chowla–Ryser theorem does not rule out the possibility of a projective plane of order 10. In 1988 C. W. H. Lam, leading a team, used a computer to prove the nonexistence of a projective plane of order 10. See C. W. H. Lam, "The search for a finite projective plane of order 10," *American Mathematical Monthly*, 98 (1991), pp. 305–318.

As we showed, there exists a projective plane of every prime power order. No projective plane of order not a prime power is known to exist, and it is conjectured that there is none.

Open problem. Construct a projective plane of order 12 or show that none exists.

A *finite affine plane* π'_n is a projective plane of order n without the ideal points and the line at infinity. A π'_n is an $(n^2, n^2 + n, n + 1, n, 1)$ BIBD. For example, a π'_3 is a (9, 12, 4, 3, 1) BIBD, which we have already seen is an $S(2, 3, 9)$ Steiner system. In general, a π'_n is an $S(2, n, n^2)$. A projective plane of order n exists if and only if there exists an affine plane of order n. See [20].

EXERCISES

6.14 An instructor has 25 students whom she wishes to divide into five groups of five students each. The students will be regrouped each class day, and the class will meet for a large number of days. How can she do the grouping so as to minimize the number of times that any two students are in the same group?

6.15 Let M be the 13 × 13 circulant matrix whose first row is the characteristic vector of the set {1, 2, 4,10}. Show that M is an incidence matrix for a (13, 4, 1) SBD, i.e., a projective plane of order 3.

6.16 Draw a projective plane of order 4.

6.17 Give a counting argument for the order of the automorphism group of a projective plane of order 3.

The automorphism group of a projective plane over a field F is a semidirect product Aut $F \cdot$ PGL(3,|F|). If F is GF(q), where $q = p^k$, then |Aut F| = k.

6.4 Latin squares

A *Latin square* L of order n is an $n \times n$ array $[L(i, j)]$ in which each row and each column contains all the elements of N_n. An $r \times n$ *Latin rectangle* consists of the first r rows of a Latin square of order n.

It is easy to create a Latin square of any order, as in the next example.

EXAMPLE 6.16 A Latin square of order 3

To create a Latin square of order 3, take the first row to be 1, 2, 3, and sucessive rows to be shifts of the first row.

1	2	3
2	3	1
3	1	2

EXAMPLE 6.17

The Cayley table of a finite group G with elements $g_1,\ldots,g_n$ yields a Latin square L of order n. Let the (i, j) entry of L be k, where $g_ig_j = g_k$. For instance, $G = \mathbf{Z}_2 \times \mathbf{Z}_2$, with elements (0, 0), (0, 1), (1, 0), (1, 1), yields the Latin square

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

Not all Latin squares come from groups. However, Latin squares are equivalent to the multiplication tables of primitive algebraic structures called quasigroups. We record some interesting definitions. A *groupoid* is a nonempty set S and a binary operation * defined on S. A *semigroup* is a groupoid in which * is an associative operation. A *monoid* is a semigroup containing a two-sided identity element e ($x * e = e * x = x$ for all x). A *group* is a monoid in which every element x in S has a two-sided inverse x^{-1} ($x * x^{-1} = x^{-1} * x = e$). A *quasigroup* is a groupoid such that given $a, b \in S$ there exist unique x, y with $a * x = b$ and $y * a = b$. A *loop* is a quasigroup containing a two-sided identity e. The literature abounds with examples of these algebraic structures. For instance, given any $S(2, 3, n)$ Steiner triple system S we may define a quasigroup whose elements are members of S by setting, for a and b distinct, $a * b = c$, where c is the unique element in a triple with a and b, and setting $a * a = a$. By adding an identity element 1 and properly extending the definition of multiplication we may turn this quasigroup into a loop. Another example of a loop is the famous Cayley loop of order 16. To see some other loops the reader should consult [7].

Suppose that we have a quasigroup with n elements, $g_1,\ldots,g_n$. By definition, each row and each column of its multiplication table is a permutation of the n elements. Therefore, replacing $g_1,\ldots,g_n$ by the numbers $1,\ldots,n$ results in a Latin square of order n. Conversely, any Latin square of order n is the multiplication table of a quasigroup of order n.

Two Latin squares L_1 and L_2 are *equivalent* if L_1 can be transformed into L_2 by the following operations:

1. Reordering rows

2. Reordering columns

3. Permuting symbols

It is an open problem to determine an asymptotic formula for the number $L(n)$ of Latin squares of order n (equivalently, the number of quasigroups of order n) and the number $L^*(n)$ of inequivalent Latin squares of order n. However, the following existence theorem allows us to formulate a lower bound for $L(n)$.

Hall's marriage theorem (1935). Let $S_1,\ldots,S_n$ be finite sets. There exist distinct $s_i \in S_i$ (for each i) if and only if the following condition holds for each k with $1 \le k \le n$: the union of any k of the S_i contains at least k elements.

The set $\{s_i\}$ is called a *system of distinct representatives* (SDR) for the S_i. If S_i is a list of men whom woman i would like to marry, then an SDR is a feasible set of marriages; hence the title of the theorem.

Proof. The necessity of the conditions is obvious. We must prove that the conditions are sufficient. Certainly, there is a representative for the first set. Assume that distinct representatives exist for the first k of the sets. We will show that an SDR can be found for $k + 1$ sets. Let T_1 be a set which has no representative assigned to it yet. If there is an element of T_1 not already occurring as a representative of one of the other k sets, then we are done. Otherwise, note that T_1 has at least one element, say t_1, and suppose that t_1 represents T_2. By hypothesis, $T_1 \cup T_2$ contains at least one element other than t_1, say t_2. If t_2 is not already a representative, then stop. If t_2 represents a set T_3, then find $t_3 \in T_1 \cup T_2 \cup T_3$. Continuing in this manner, we find a collection $\{t_i\}$ such that $t_i \in T_1 \cup \cdots \cup T_i$ and t_i represents T_{i+1} (for $i < a$), and t_a is not a representative yet. Now we change some representatives by pairing t_a with a set T_a' where $a \le a'$. This process continues until T_1 is paired with a representative. These new pairings, together with the unchanged pairings, constitute an SDR for $k + 1$ sets.

The following corollary is proved in [12].

Corollary. If $S_1,\ldots,S_n$ are sets possessing an SDR, and if the smallest set has size $t < n$, then the S_i possess at least $t!$ SDRs.

Lower bound for the number of Latin squares. $L(n) \ge n!(n-1)! \ldots 2!1!$.

Proof. We will show that for each r, where $1 \le r \le n - 1$, an $r \times n$ Latin rectangle may be extended to an $(r + 1) \times n$ Latin rectangle in at least $(n - r)!$ ways. Given an $r \times n$ Latin rectangle, let S_i be the set of numbers not yet used in column i. Clearly, an SDR could be used as the $(r + 1)$st row of the Latin square. Now, each element m, with $1 \le m \le n$, has occurred in r rows and hence in r columns of the Latin rectangle thus far. Therefore, each element occurs in exactly $n - r$ of the S_i. For each k, the union of k of the S_i contains $k(n - r)$ elements (counting repetitions). As each element occurs in at most $n - r$ of these S_i, the union must contain at least k elements, and the criterion in Hall's theorem is satisfied. Hence, there is an SDR for the S_i.

Because each S_i has size $n - r$, the corollary guarantees the existence of at least $(n - r)!$ SDRs. The inequality on $L(n)$ is established by applying the above estimate as each successive row is added to the Latin square.

By a permutation of its rows and columns, any Latin square may be written with $1,\ldots,n$ as its first row and first column. Such a Latin square is said to be *standardized*. If $L'(n)$ is the number of inequivalent standardized Latin squares of order n, then $L(n) = n!(n-1)!L'(n)$. Table 6.1 gives the values of $L'(n)$ for $1 \le n \le 7$.

Table 6.1 The number of standardized Latin squares.

n	1	2	3	4	5	6	7
$L'(n)$	1	1	1	4	56	9408	16942080

Open problem. Find a formula for $L'(n)$.

EXERCISES

6.18 Verify that $L'(4) = 4$.

6.19 Use a computer to verify that $L'(5) = 56$.

6.5 MOLS and OODs

Two Latin squares $L_1 = [L_1(i, j)]$ and $L_2 = [L_2(i, j)]$ of order n are *orthogonal* if, for every $(a, b) \in \mathbf{N}_n \times \mathbf{N}_n$, there is an ordered pair (i, j) with $(L_1(i, j), L_2(i, j)) = (a, b)$. In other words, the ordered pairs $(L_1(i, j), L_2(i, j))$ take each of the n^2 values in $\mathbf{N}_n \times \mathbf{N}_n$ exactly once. The two Latin squares of Figure 6.2 are orthogonal.

Figure 6.2 Two orthogonal Latin squares of order 3.

1	2	3		1	2	3
3	1	2		2	3	1
2	3	1		3	1	2

A set of *mutually orthogonal Latin squares*, or *MOLS*, is a set in which every pair is orthogonal. MOLS are also called *pairwise orthogonal* Latin squares. We define $m(n)$ to be the maximum possible number of MOLS of order n. Leonhard Euler (1707–1783) introduced the ideas of Latin squares and MOLS in 1782 when he asked whether there are two MOLS of order 6. He believed the answer is no and therefore conjectured that $m(6) = 1$. This was proved by G. Tarry in 1900. Euler also conjectured that $m(n) = 1$ whenever $n \equiv 2 \pmod 4$, but it was shown in 1960 by R. C. Bose, E. T. Parker, and S. S. Shrikhande that $m(n) \geq 2$ except when $n = 1$, 2, or 6. For example, there are two MOLS of order 10. However, it is not known whether there are three MOLS of order 10.

Theorem. For all $n \geq 2$, we have $m(n) \geq n - 1$.

Proof. Suppose that there is a set of n MOLS of order n. By a permutation of symbols, the first row of each Latin square can be changed to $1,\ldots,n$, and permuting symbols clearly does not disturb orthogonality. Now, by the pigeonhole principle, since none of the (2, 1) entries of the n MOLS can equal 1, some two Latin squares have (2, 1) entry equal to i, with $2 \leq i \leq n$. But these Latin squares are not orthogonal because the ordered pair (i, i) occurs twice in the list of ordered pairs of entries.

When $n = p^k$ for a prime p, we can construct $n - 1$ MOLS of order n. Suppose that F is the field GF(p^k), and let $F = \{0 = f_0,\ldots,f_{n-1}\}$. For each m, where $1 \leq m \leq n - 1$, define the Latin square $L_m = [L_m(i, j)]_{n \times n}$, where $0 \leq i, j \leq n - 1$, by $L_m(i, j) = f_m f_i + f_j$. It is a simple matter to check that each L_m is a Latin square. To check the orthogonality condition, observe that $(f_m f_i + f_j, f_n f_i + f_j) =$ =

$(f_m f_k + f_l, f_n f_k + f_l)$ implies that $f_i = f_k$ and $f_j = f_l$, so all the ordered pairs are distinct.

For example, starting with the three-element field $\{0, 1, 2\}$, we produce the two MOLS

$$\begin{array}{lllllll} 0\cdot1+0 & 0\cdot1+1 & 0\cdot1+2 & & 0\cdot2+0 & 0\cdot2+1 & 0\cdot2+2 \\ 1\cdot1+0 & 1\cdot1+1 & 1\cdot1+2 & & 1\cdot2+0 & 1\cdot2+1 & 1\cdot2+2 \\ 2\cdot1+0 & 2\cdot1+1 & 2\cdot1+2 & & 2\cdot2+0 & 2\cdot2+1 & 2\cdot2+2 \end{array}$$

or

$$\begin{array}{lll} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{array} \qquad \begin{array}{lll} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{array}$$

(two orthogonal Latin squares equivalent to those in Figure 6.2).

We have described how a projective plane of order n can be constructed from the field GF(n). The above argument shows that $n - 1$ MOLS of order n can be constructed from GF(n). In fact, a projective plane of order n is equivalent to a set of $n - 1$ MOLS of order n.

Theorem. A set of $n - 1$ MOLS of order n is equivalent to a projective plane of order n.

Proof. Suppose that we are given $n - 1$ MOLS: $L_1, \ldots, L_{n-1}$. Let

$$S = \{(i, j) \in \mathbf{N}_n \times \mathbf{N}_n\} \cup \{i : i \in \mathbf{N}_{n-1}\} \cup \{0, \infty\},$$

and define

$$\begin{aligned} l_{\infty,k} &= \{(k, i) : i \in \mathbf{N}_n\} \cup \{\infty\}, \quad 1 \le k \le n \\ l_{0,k} &= \{(i, k) : i \in \mathbf{N}_n\} \cup \{0\}, \quad 1 \le k \le n \\ l_{x,y} &= \{(i, j) : L_x(i, j) = y\} \cup \{x\}, \quad 1 \le x \le n-1,\ 1 \le y \le n \\ l_{\infty} &= \{i : i \in \mathbf{N}_{n-1}\} \cup \{0, \infty\}. \end{aligned}$$

We leave it to the reader to check that the incidence matrix for the set of points S and the lines $l_{\infty,k}$, $l_{0,k}$, $l_{x,y}$, l_∞ is an $(n^2 + n + 1, n + 1, 1)$ SBD.

Reversing the above construction completes the equivalence.

The reader may find it instructive to apply the construction technique to the two MOLS of Figure 6.2 to construct a projective plane of order 3. It will be equivalent to the one of Figure 6.1.

Let us extend the definition of orthogonality to any two square matrices (not necessarily Latin squares). We say that two $n \times n$ matrices A and B with entries from $\mathbf{N}_n$ are orthogonal if the ordered pairs $(A(i, j), B(i, j))$ take each of the n^2 values in $\mathbf{N}_n \times \mathbf{N}_n$ exactly once. Observe that the matrices

$$R = \begin{bmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & 2 & 3 & \ldots & n \\ \vdots & & & & \\ 1 & 2 & 3 & \ldots & n \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 2 & 2 & 2 & \ldots & 2 \\ \vdots & & & & \\ n & n & n & \ldots & n \end{bmatrix}$$

are orthogonal. With this generalized definition of orthogonality, we can give an elegant characterization of Latin squares: an $n \times n$ matrix is an order n Latin square if and only if it is orthogonal to both R and C. Thus, any k MOLS of order n are part of a family of $k + 2$ mutually orthogonal matrices. Conversely, any $k + 2$ mutually orthogonal $n \times$ matrices may be transformed into R, C, and k MOLS of order n. For if a matrix M is orthogonal to another matrix, then M contains each

of the numbers 1,…,n exactly n times. Therefore, choosing two matrices M and N from the set of k + 2 orthogonal matrices, we may transform M into R and N into C by a simultaneous permutation of the entries of all the matrices. Discarding R and C, we are left with k MOLS of order n.

This characterization replaces the notion of Latinicity with the more essential notion of orthogonality. Accordingly, we define an *ordered orthogonal design* of order n and depth s (an (n, s) OOD) to be an $s \times n^2$ matrix (m_{ij}) with entries 1,…,n such that every two rows are orthogonal. That is, for every pair of rows u and v, every ordered pair (a, b) with $1 \leq a, b \leq n$ occurs exactly once among the ordered pairs (m_{ui}, m_{vi}). For example, Figure 6.3 shows a (3, 4) OOD derived from the two Latin squares of order 3 in Figure 6.2. An OOD is also called an OA (orthogonal array).

Figure 6.3 A (3,4) OOD.

1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3
1	2	3	3	1	2	2	3	1
1	2	3	2	3	1	3	1	2

The design inherent in the theorem can be produced directly from an $(n, n + 1)$ OOD. In general, an *OOD-net* based on an (n, s) OOD is a collection of m points corresponding to the columns of the OOD and s *pencils* of parallel lines, where point x_j is on line y of the ith pencil if the (i, j) entry of the matrix is y. The OOD-net of an $(n, n + 1)$ OOD is an affine plane of order n which may be extended to a projective plane of order n by the addition of one ideal point for each row of the OOD and an ideal line.

In summary, we have shown the equivalence of the following discrete configurations:

- a projective plane of order n;
- a collection of $n - 1$ MOLS of order n;
- an $(n, n + 1)$ OOD.

Open problem. Determine the values of n for which these structures exist.

A possible conjecture consistent with what is known is that these structures exist if and only if n is a prime power.

EXERCISES

6.20 Prove that the $n \times n$ arrays $A = [a_{ij}]$ and $B = [b_{ij}]$, where $a_{ij} = i + j$ and $b_{ij} = i - j$ (modulo n) are orthogonal, when n is an odd number.

6.21 Construct a set of three 4×4 MOLS (equivalent to a projective plane of order 4).

6.22 Construct a set of four 5×5 MOLS (equivalent to a projective plane of order 5).

6.23 Construct a (5, 6) OOD.

6.6 Hadamard matrices

In 1893 Jacques Hadamard considered a basic problem about the maximum absolute value of the

determinant of a matrix with bounded entries.

Hadamard's theorem (1893). Suppose $A = [a_{ij}]$ is matrix of order m with $-1 \le a_{ij} \le 1$ for all i and j. Then $|\det A| \le m^{m/2}$, and the upper bound is obtained if and only if $a_{ij} = \pm 1$ for all i and j and $AA^t = mI$.

Proof. The rows of A are vectors in $\mathbf{R}^n$ of length at most $m^{1/2}$, and they span a parallelepiped of volume $|\det A|$. This volume is clearly maximized when the vectors are mutually orthogonal and of maximum possible length, and in this case the volume is the product of the lengths, $m^{m/2}$.

A matrix $A = [a_{ij}]_{m \times m}$ with $a_{ij} = \pm 1$ and $AA^t = mI$ is called a *Hadamard matrix* of order m. The condition $AA^t = mI$ means that the dot product of any two distinct rows of A is zero. (The same is true for columns, as $A^tA = A^{-1}\ (AA^t)A = A^{-1}(mI)A = mI$.) Therefore, without regard to the volume argument given in the proof above, if A is a Hadamard matrix of order m, then $m^m = \det AA^t = (\det A)^2$, which implies that $|\det A| = m^{m/2}$.

Notice that the theorem does not address the question of the maximum value of $|\det A|$ when there is no Hadamard matrix of order m. However, $|\det A|$ does attain a maximum, as it is a continuous function defined on a compact set (the cube $[-1, 1]m^2$). Furthermore, because $\det A$ is a linear function of each entry a_{ij} (i.e., a straight line), the function $y = |\det A|$ is concave upward and therefore the maximum of $|\det A|$ occurs when each $a_{ij} = -1$ or 1. The maximum determinant may also occur for other matrices, in case the coefficient of the first-order term in the linear equation just described is 0. For example,

$$\begin{vmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \\ -1 & 1 & 1 \end{vmatrix} = 4.$$

The *Kronecker product* $A \otimes B$ of two square matrices $A = [a_{ij}]_{m_1 \times m_1}$ and $B = [b_{ij}]_{m_2 \times m_2}$ is the square matrix $A \otimes B = [a_{ij}B]_{m_1 m_2 \times m_1 m_2}$. The Kronecker product produces larger Hadamard matrices from smaller ones. For example, Figure 6.4 shows Hadamard matrices A and B of orders 2 and 4, respectively, with $B = A \otimes A$.

Figure 6.4 Hadamard matrices of orders 2 and 4.

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

If H is a Hadamard matrix, then so is

$$\begin{bmatrix} H & H \\ -H & H \end{bmatrix}.$$

Theorem. There exists a Hadamard matrix of order $m = 2^k$, where k is any positive integer.

If A is a Hadamard matrix, then any permutation of the rows or columns of A is a Hadamard matrix. Also, any row or column of A may be multiplied by –1 with the result still a Hadamard matrix. With these operations it is possible to alter any Hadamard matrix so that its first row and first column consist of all 1′s. Such a Hadamard matrix is said to be *normalized*.

Theorem. If A is a Hadamard matrix of order $m > 2$, then m is a multiple of 4.

Proof. Normalize A and permute its columns so that its first three rows look like this:

$$\begin{array}{cccccccccccc}
1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 & \dots \\
1 & 1 & \dots & 1 & 1 & \dots & -1 & -1 & \dots & -1 & -1 & \dots \\
1 & 1 & \dots & -1 & -1 & \dots & 1 & 1 & \dots & -1 & -1 & \dots \\
 & & & \underbrace{}_{a} & & & \underbrace{}_{b} & & & \underbrace{}_{c} & & \underbrace{}_{d}
\end{array}$$

(The variables a, b, c, d are to be determined.) Four equations are immediate from the definition of a Hadamard matrix:

$$\begin{aligned}
a+b+c+d &= m \\
a+b-c-d &= 0 \\
a-b+c-d &= 0 \\
a-b-c+d &= 0.
\end{aligned}$$

Adding the equations yields $4a = m$, from which it follows that $a = b = c = d = m/4$. Therefore, m is a multiple of 4, and furthermore we have found that any row after the first has $m/2$ 1′s and $m/2$ –1′s, and any two rows (not including the first) have 1′s together in $m/4$ columns.

It is conjectured that the condition of the theorem is sufficient as well as necessary.

Conjecture. There exists a Hadamard matrix of every order a multiple of 4.

The smallest order for which the existence of a Hadamard matrix is not certain is 668.

Open problem. Determine whether there is a Hadamard matrix of order 668.

Theorem. Let H be a normalized Hadamard matrix of order $4n \geq 8$. Deleting the first row and column of H and changing each –1 to a 0 results in an incidence matrix of a $(4n - 1, 2n - 1, n - 1)$ SBD. Conversely, starting with an incidence matrix of a $(4n - 1, 2n - 1, n - 1)$ SBD, changing each 0 to –1, and adding a first row and first column of all 1′s, yields a normalized Hadamard matrix of order $4n$.

Proof. Let X be the submatrix of H formed by deleting its first row and column. The Hadamard conditions imply that $XJ = JX = -J$ and $XX^t = 4nI - J$. When each –1 is switched to 0 a new matrix $Y = \frac{1}{2}(X + J)$ results. We check that Y is the incidence matrix of a $(4n - 1, 2n - 1, n - 1)$ SBD: $JY = YJ = \frac{1}{2}(-J + (4n - 1)J) = (2n - 1)J$ and $YY^t = \frac{1}{4}(X + J)(X^t + J) = nI + (n - 1)J$. The proof of the reverse construction is similar.

EXAMPLE 6.18

The (7, 3, 1) SBD is equivalent to a Hadamard matrix of order 8.

A $(4n - 1, 2n - 1, n - 1)$ SBD created this way is called a *Hadamard design* of order n. A 3-$(4n, 2n, n - 1)$ design may be formed by taking complements of a Hadamard design H together with the blocks of H joined to a new point ∞.

Hadamard designs and projective planes are two extreme types of (v, k, λ) designs. For if a (v, k, λ) design exists, then

$$4n - 1 \leq v \leq n^2 + n + 1, \tag{6.7}$$

where $n = k - \lambda$. To prove this inequality, we let $\lambda' = v - 2k + \lambda$. Then $\lambda + \lambda' = v - 2n$ and $\lambda\lambda' = n(n - 1)$. The upper bound follows from the observation that $\lambda' \geq 1$, and the lower bound from the arithmetic mean–geometric mean inequality, $(\lambda + \lambda')^2 \geq 4\lambda\lambda'$. One can show that the upper bound is met if and only if the design is a projective plane of order n (or its complement) and the lower bound is met only for a Hadamard design of order n (or its complement).

Let H be a normalized Hadamard matrix of order $m = 4n$, with each -1 changed to 0, and let A be the code consisting of the rows of H and the binary complements of these rows. Clearly, A is a code of dimension $4n$ containing $8n$ codewords. We claim that the distance of A is $2n$. We are guaranteed that any two rows of H disagree in exactly $2n$ places. Therefore, any two rows of the binary complementary matrix H' disagree in $2n$ entries. Suppose that $a \in H$ and $b \in H'$. If $a = b^c$, then $d(a, b) = m$. If not, then $d(a, b)$ equals the number of components in which a and b^c agree, which is $2n$. This $(4n, 8n, 2n)$ code is called a *Hadamard code*. It is capable of detecting $2n - 1$ errors and correcting $n - 1$ errors.

EXAMPLE 6.19

The (7, 3, 1) SBD yields an (8, 16, 4) code. We saw this code in Chapter 5.

EXAMPLE 6.20

From the complement of the quadratic residues construction for $p = 11$, we obtain the incidence matrix of an (11, 5, 2) SBD:

$$Y = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

From Y we construct a normalized Hadamard matrix of order 12. Thus

$$H = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\
1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \\
1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\
1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1
\end{bmatrix}.$$

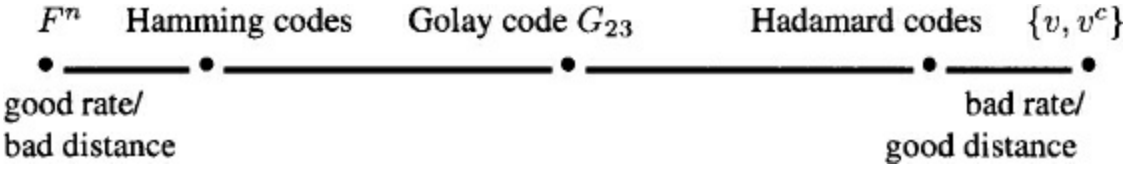
We shall see in the next section that Y is one of the main ingredients in producing a perfect $(23, 2^{12}, 7)$ code. Switching -1 back to 0, the rows of H and their complements constitute a (12, 24, 6) code which detects five errors and corrects two errors. The words of this code are listed below.

1 1 1 1 1 1 1 1 1 1 1 1
1 0 0 1 0 0 0 1 1 1 0 1
1 1 0 0 1 0 0 0 1 1 1 0
1 0 1 0 0 1 0 0 0 1 1 1
1 1 0 1 0 0 1 0 0 0 1 1
1 1 1 0 1 0 0 1 0 0 0 1
1 1 1 1 0 1 0 0 1 0 0 0
1 0 1 1 1 0 1 0 0 1 0 0
1 0 0 1 1 1 0 1 0 0 1 0
1 0 0 0 1 1 1 0 1 0 0 1
1 1 0 0 0 1 1 1 0 1 0 0
1 0 1 0 0 0 1 1 1 0 1 0

0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 0 1 1 1 0 0 0 1 0
0 0 1 1 0 1 1 1 0 0 0 1
0 1 0 1 1 0 1 1 1 0 0 0
0 0 1 0 1 1 0 1 1 1 0 0
0 0 0 1 0 1 1 0 1 1 1 0
0 0 0 0 1 0 1 1 0 1 1 1
0 1 0 0 0 1 0 1 1 0 1 1
0 1 1 0 0 0 1 0 1 1 0 1
0 1 1 1 0 0 0 1 0 1 1 0
0 0 1 1 1 0 0 0 1 0 1 1
0 1 0 1 1 1 0 0 0 1 0 1

The families of codes we have encountered can be arranged on a continuum from good rate/bad distance to bad rate/good distance, as in Figure 6.5. At the left extreme is the code F^n, in which every vector is a codeword. The rate is 1, but the code is incapable of correcting any errors. At the right extreme is a code consisting of any vector v and its complement v^c. Although this code has the highest possible distance, $d = n$, its rate is $1/n$, the lowest possible. The family of Hamming codes are capable of correcting $e = 1$ error, and the rates tend to 1. Therefore, the Hamming codes converge to the left endpoint of the continuum. For the Hadamard codes,

Figure 6.5 The world of codes.



$$r(A) = \frac{\log_2 8n}{4n} = \frac{3 + \log_2 n}{4n} \to 0 \text{ (as } n \to \infty). \tag{6.8}$$

Since their rates converge to 0, the Hadamard codes converge to the right endpoint of the continuum.

The $(23, 2^{12}, 7)$ Golay code G_{23}, which we will construct in the next section, has rate 12/23 and corrects $e = 3$ errors.

We conclude the discussion of combinatorial designs by constructing three large, interesting, related configurations, namely, the $(23, 2^{12}, 7)$ Golay code G_{23} (the only perfect multi-error-correcting binary code), the $S(5, 8, 24)$ Steiner system (consisting of the weight 8 codewords of the extended Golay code G_{24}), and Leech's lattice $\mathcal{L}$ (a 24-dimensional lattice obtained from G_{24} which generates a surprisingly dense sphere packing).

EXERCISES

6.24 Construct a Hadamard matrix of order 8. Change this into a (7, 3, 1) SBD, i.e., a Fano Configuration. Show that the code produced from this matrix has parameters (8, 16, 4).

6.25 Show that the determinant of a square matrix is a linear function of each of its entries. That is, if $A = [a_{ij}]_{m \times m}$ and a_{ij} is a fixed number for each (i, j) except (i_0, j_0), then there exist numbers α and β, not depending on $a_{i_0 j_0}$, such that $\det A = \alpha a_{i_0 j_0} + \beta$.

6.26 Use the Kronecker product to construct a Hadamard matrix of order 16. Construct a (15, 7, 3) SBD. What code does this design give?

6.27 Construct a (19, 9, 4) SBD. What code does this design give?

6.7 The Golay code and $S(5, 8, 24)$

We have indicated that the only feasible parameters for perfect binary codes with $e > 1$ are $(90, 2^{78}, 5)$ and $(23, 2^{12}, 7)$. An exercise calls for a proof that no $(90, 2^{78}, 5)$ code exists. We will construct a $(23, 2^{12}, 7)$ code called the *Golay code,* G_{23}, which was discovered by Marcel Golay (1902–1989). It is a perfect 3-error correcting code with 2^{12} words, sitting inside F^{23}. As a bonus, we will find that

certain words in the extended Golay code G_{24} constitute a Steiner system $S(5, 8, 24)$.

Let M be an (11, 6, 3) SBD based on quadratic residues modulo 11. Let G be the 12×24 matrix

$$\left[\begin{array}{cccccccccccc|c|cccccccccccc} 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & I_{11} & & & & & & & 0 & & & & & M & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 1 & & & & & & & & & & & & 0 & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right],$$

where I_{11} is the order 11 identity matrix.

The matrix G defines a linear transformation $G: F^{12} \to F^{24}$, with $x \mapsto xG$. The (linear) code G_{24} is the image $\{xG : x \in F^{12}\}$. We call G a *generating matrix* for G_{24}. It is easy to use a computer to produce the codewords of G_{24} and thereby verify that it is a $(24, 2^{12}, 8)$ code. But we can do so without a computer as follows.

We leave it to the reader to use elementary row operations to reduce G to row echelon form and thus show that G has rank 12. It follows that G_{24} has parameters $(24, 2^{12}, d)$, where d is to be determined. We proceed to find the weight distribution of G_{24}.

Recall from Exercise 5.3 that if x and y are any two vectors of the same length, then $w(x + y) = w(x) + w(y) - 2x \cdot y$.

If r_i and r_j are different rows of G, then $r_i \cdot r_j$ is even. This follows from the row dot product property of M.

Now we show that if $x \in G_{24}$, then $w(x)$ is a multiple of 4. Any codeword is a linear combination (over F) of the rows of G, so we can write $x = r_1 + r_2 + \cdots r_n$ (with relabeling of rows). We use induction on n. If $n = 1$, then by inspection $w(x)$ is a multiple of 4 (the last row has weight 12 and every other row has weight 8). Now if $x = r_1 + r_2 + \cdots + r_n\, r_{n+1}$ then

$$w(x) = w(r_1 + \cdots + r_n) + w(r_{n+1}) - 2(r_1 + \cdots + r_n) \cdot r_{n+1},$$

which, by remarks made earlier, is a multiple of 4. This completes the induction.

Therefore, the possible weights of words of G_{24} are 0, 4, 8, 12, 16, 20, 24. Clearly, $0 \in G_{24}$ and $w(0) = 0$. Also, $r_1 + r_2 + \cdots + r_{24} = (1,\ldots,1)$, so there is a word of weight 24. It follows that the binary complement of any codeword is also a codeword and therefore the weight distribution of G_{24} is symmetric. This distribution, as we have found so far, is given by Table 6.2. The variables α, β, γ have yet to be determined.

Table 6.2 The partially determined weight distribution of G_{24}.

weight	0	4	8	12	16	20	24
number of words	1	α	β	γ	β	α	1

Suppose that $x \in G_{24}$. Let $L(x)$ be the left-hand string of length 12 of x and $R(x)$ the right-hand string. We can now represent x as $x = [L(x), R(x)]$. We claim that $[R(x), L(x)] \in G_{24}$; that is, G_{24} is invariant under the permutation of coordinates

$$\tau = (1\ 13)(2\ 14)(3\ 15)\ldots(12\ 24).$$

We say that G_{24} is *self-symmetric*. Let v' denote the vector obtained from v by switching the right and left halves. We leave it to the reader to show that for each row r of G, the vector r' is a linear combination of the rows r_i. It follows that G_{24} is invariant under τ, as every codeword of G_{24} is a sum of rows r_i.

Next we observe that $w(L(x))$ is even whenever $x \in G_{24}$, because the sum of k rows when k is even yields $w(L(x)) = k$ and the sum when k is odd yields $w(L(x)) = k + 1$.

Because $w(L(x)) + w(R(x))$ is a multiple of 4, it follows that $w(R(x))$ is always even.

Finally, we will show that $d \geq 8$, from which it follows that $d = 8$, because the first row of G has weight 8. We need only show that no codeword has weight 4. If x has weight 4, then $(w(L(x)), w(R(x))) = (0, 4), (4, 0)$, or $(2, 2)$. If $w(L(x)) = 0$, then x must be r_{12}, but then $w(R(x)) = 8 > 4$. Because G_{24} is self-symmetric, the (4, 0) case is ruled out also. For the (2, 2) case, we can sum any one or two of the first 11 rows of G and then add or not add the 12th row. In each case the resulting codeword has weight 6, not 2.

Therefore, G_{24} is a $(24, 2^{12}, 8)$ code. Deleting any coordinate of G_{24} produces the Golay code G_{23}, a code with parameters $(23, 2^{12}, 7)$.

Theorem. The Golay code G_{23} is a (perfect) $(23, 2^{12}, 7)$ code.

Let us complete the weight distribution table of G_{24}. We know that G_{24} has no codewords of weight 4 or 20. How many words have weight 8? If $w(x) = 8$, then $(w(L(x)), w(R(x)))$ equals $(0, 8), (2, 6), (4, 4), (6, 2)$, or $(8, 0)$. A glance at G shows that $(0, 8)$ is impossible, and hence by self-symmetry $(8, 0)$ is impossible. To obtain a weight partition $(2, 6)$, we can add one or two rows of the first 11 rows of G and then add or not add the 12th row. There are $2(\binom{11}{1} + \binom{11}{2}) = 132$ possibilities. Likewise, there are 132 ways to obtain a codeword with weight partition $(6, 2)$. To get a $(4, 4)$ weight partition, we add either three or four rows of G. The number of choices is $\binom{11}{3} + \binom{11}{4} = 495$. Altogether, the number of words of weight 8 in G_{24} is $132 + 132 + 495 = 759$.

We display the complete weight distribution of G_{24} in Table 6.3.

Table 6.3 The weight distribution of G_{24}.

weight	0	8	12	16	24
number of words	1	759	2576	759	1

We are now ready to find the Steiner system $S(5, 8, 24)$ sitting inside G_{24}. From the parameter theorem for $S(5, 8, 24)$, we obtain $\lambda_5 = 1$, $\lambda_4 = 5$, $\lambda_3 = 21$, $\lambda_2 = 77$, $\lambda_1 = 253$, and $\lambda_0 = 759$. As λ_0 is the

number of blocks, it is clear that the words of weight 8 in G_{24} should form the blocks of $S(5, 8, 24)$. Let S be the set of coordinates 1,…,24 and let C be the collection of sets of eight coordinates which equal 1 in codewords of weight 8 in G_{24}. We need to check that the $t = 5$ and $\lambda = 1$ conditions are met. This means that every five-element subset of S is contained in exactly one block in C. Equivalently, every vector of weight 5 is covered by exactly one codeword of weight 8 in G_{24}.

A vector v of weight 5 cannot be covered by two different codewords x and y of weight 8, or else $w(x + y) \leq 6$, a contradiction.

Therefore, it remains to demonstrate that there are enough codewords of weight 8 to satisfy the $\lambda = 1$ condition. There are $\binom{24}{5}$ vectors of weight 5, and each of the 759 codewords of weight 8 covers $\binom{8}{5}$ of them. A simple calculation shows that $\binom{24}{5} = 759\binom{8}{5}$. Therefore, S and C make up the desired Steiner system $S(5, 8, 24)$.

Theorem. The weight 8 codewords of G_{24} are an $S(5, 8, 24)$ Steiner system.

EXERCISES

6.28 Show that the weight distribution of G_{23} is:

weight	0	7	8	11	12	15	16	23
number of words	1	253	506	1288	1288	506	253	1

6.29 Use the generating matrix G and a computer to find the codewords and weight distribution of the $(24, 2^{12}, 8)$ code.

6.30 Find the parity check matrix of the $(24, 2^{12}, 8)$ code.

6.31 Show that a Steiner system $S(5, 6, 12)$ can be constructed by fixing a codeword x of weight 12 in G_{24} and taking all codewords which intersect x in six places.

6.32 Show that M_{12}, the automorphism group of the above Steiner system, has $P(12, 5) = 12!/7!$ elements.

6.8 Lattices and sphere packings

An *n-dimensional lattice* $\mathcal{L}$ is a subset of $\mathbf{R}^n$ such that:

1. $0 = (0,\ldots,0) \in \mathcal{L}$;

2. $x \in \mathcal{L}$ implies that $-x \in \mathcal{L}$;

3. $x, y \in \mathcal{L}$ imply that $x + y \in \mathcal{L}$.

We assume that $\mathcal{L}$ contains a point other than the origin (hence $\mathcal{L}$ is infinite) and that $\mathcal{L}$ is *discrete*, which means that $\mathcal{L}$ has a finite intersection with any compact subset of $\mathbf{R}^n$.

The *Euclidean distance* between two points x and y in $\mathbf{R}^n$ is

$$d(x, y) = ((x_1 - y_1)^2 + \cdots + (x_n - y_n)^2)^{1/2}, \tag{6.9}$$

and the *norm* of x is

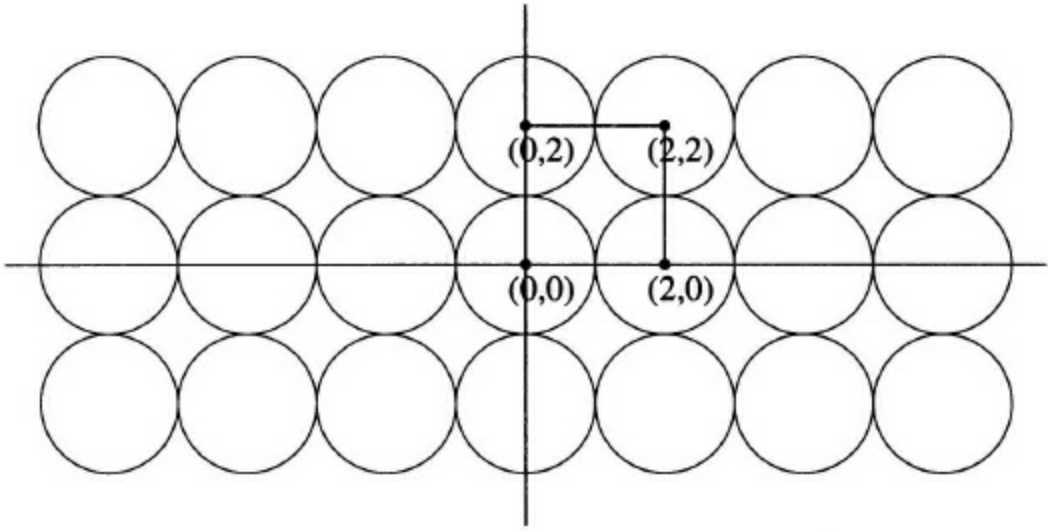
$$\|x\| = d(x, 0) = (x_1^2 + \cdots + x_n^2)^{1/2}. \tag{6.10}$$

If $\mathcal{L}$ is a discrete n-dimensional lattice containing a non-origin point x, then the *Euclidean sphere* $\{y \in \mathbf{R}^n : 0 \leq \|y\| \leq \|x\|\}$ contains only finitely many points of $\mathcal{L}$. The minimum positive distance between 0 and a point in this intersection is the *minimum distance* of $\mathcal{L}$, denoted $d(\mathcal{L})$. Two lattice points are *neighbors* if they are separated by this minimum distance. Because a lattice is clearly translation invariant, any base point determines the same minimum distance $d(\mathcal{L})$ to a neighbor. If spheres of radius $\frac{1}{2}d(\mathcal{L})$ are centered at each lattice point, these spheres touch only at their surfaces. Such a placement of spheres is called a *lattice sphere packing* of $\mathbf{R}^n$.

It is desirable to know how densely spheres may be packed in $\mathbf{R}^n$ via a lattice packing or otherwise. Related to this question is the computation of the maximum number of spheres touching a given sphere in a lattice packing. Specifically, let $c(\mathcal{L})$, the *contact number* of $\mathcal{L}$, be the number of spheres of radius $\frac{1}{2}d(L)$ which touch the sphere centered at the origin. Equivalently, $c(\mathcal{L})$ is the number of lattice points at distance $d(\mathcal{L})$ from 0.

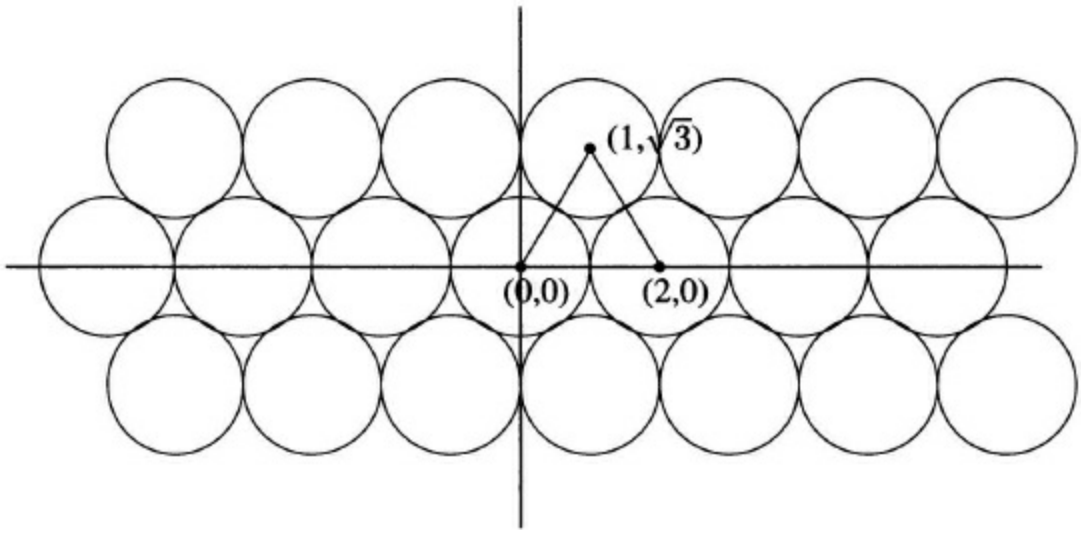
Figure 6.6 shows part of a two-dimensional lattice with distance 2 and contact number 4. This lattice is called $(2\mathbf{Z})^2$, as it consists of all points in the plane with even integer coordinates. In general, the n-dimensional lattice $(2\mathbf{Z})^n$ consists of all points in $\mathbf{R}^n$ with even integer coordinates.

Figure 6.6 A lattice packing with contact number 4 and density $\doteq 0.79$.



We can easily calculate the amount of the plane enclosed by the circles. Consider a square of area 4 whose vertices are four lattice points (as indicated in the figure). Such a square is called a *fundamental region* of the packing, as this region repeats periodically to give the complete pattern of the packing. Because the square contains four circular quadrants whose total area is π, we say that the *density* of the packing is $\pi/4 \doteq 0.79$. However, the densest possible packing in $\mathbf{R}^2$ is not based on the lattice $(2\mathbf{Z})^2$, but on the *triangular lattice* T shown in Figure 6.7. The fundamental region of T is an equilateral triangle of side length 2 and area $\sqrt{3}$. As this triangle contains three sixths of a unit circle, the density of the packing is $\pi/(2\sqrt{3}) \doteq = 0.90$. The contact number, 6, is also greater for T than for $(2\mathbf{Z})^2$. Because a unit sphere in the $(2\mathbf{Z})^n$ packing has two neighbors in each of n directions, this lattice has contact number $2n$. We shall see that the contact number can be greatly improved in 24 dimensions.

Figure 6.7 A lattice packing with contact number 6 and density $\doteq 0.90$.



Open problem. Find, for each k, the highest possible contact number of a lattice sphere packing in $\mathbf{R}^k$.

EXERCISES

6.33 Show that the contact number of $\mathbf{Z}^n$ is $2n$.

6.34 What is the maximum possible contact number $c(\mathcal{L})$ for a lattice packing $\mathcal{L}$ in $\mathbf{R}^3$? What is the density of this packing?

6.9 Leech's lattice

In 1965 John Leech (1926–1992) proved that the extended Golay code G_{24} can be used to construct a 24-dimensional lattice with a remarkably high contact number. Leech's lattice $\mathcal{L}$ is defined as follows. For each codeword $c = (c_1,\ldots,c_{24}) \in G_{24}$ and each integer m, let $\mathcal{L}(c, m)$ be the set of integer 24-tuples $(x_1,\ldots,x_{24})$ for which

$$\text{(A)} \quad \sum_{i=1}^{24} x_i = 4m \quad \text{and}$$

$$\text{(B)} \quad x_i \equiv \begin{cases} m & \pmod 4 \quad \text{if } c_i = 0 \\ m+2 & \pmod 4 \quad \text{if } c_i = 1. \end{cases}$$

Leech's lattice $\mathcal{L}$ is the union of all the $\mathcal{L}(c, m)$.

We need to verify that $\mathcal{L}$ is a lattice. Because the all-0 string is a codeword in G_{24}, it follows that $0 \in \mathcal{L}$ (with $m = 0$). If $x \in \mathcal{L}$, then $x \in \mathcal{L}(c, m)$ for some $c \in G_{24}$ and some integer m. It is easy to show that $-x \in \mathcal{L}(c, -m)$, and hence $-x \in L$. We will show that if $x, y \in \mathcal{L}$ with $x \in \mathcal{L}(c, m)$ and $y \in \mathcal{L}(d, n)$, then $x+y \in \mathcal{L}(c+d, m+n)$. Clearly, $\sum(x_i+y_i) = \sum x_i + \sum y_i = 4m+4n = 4(m+n)$, so condition (A) is satisfied. As for condition (B), if c_i and d_i are of opposite parity, then $c_i + d_i = 1$ and $x_i y_i \equiv m + n + 2 \pmod 4$. If c_i and d_i have the same parity, then $c_i + d_i = 0$ and $x_i + y_i = m + n \pmod 4$. Therefore, condition (B) is satisfied and $\mathcal{L}$ is a lattice.

So far, we have not used any particular property of the Golay code other than its linearity.

We now calculate the distance $d(\mathcal{L})$ of Leech's lattice by finding the smallest value of

$\|x\| = (x_1^2 + \cdots + x_{24}^2)^{1/2}$ for a lattice point other than the origin. Condition (B) implies that all the x_i are even (if m is even) or all the x_i are odd (if m is odd). If all the x_i are odd and some x_i satisfies $|x_i| \geq 3$, then $\sum x_i^2 \geq 23(1) + 3^2(1) = 32$. We shall soon see than 32 is, in fact, the minimum value of $\|x\|^2$. Recall that the codewords of G_{24} have weights 0, 8, 12, 16, and 24. We say that they have *shapes* 0^{24}, $0^{16}1^8$, $0^{12}1^{12}$, 0^81^{16}, and 1^{24}. If $|x_i| = 1$ for all x_i, then x has shape $(+1)^a\ (-1)^b$ where (a, b) is one of (24, 0), (16, 8), (12, 12), (8, 16), (0, 24). Therefore $\Sigma\ x_i = 24$, 8, 0, –8, or –24. In each case, the sum is $4m$ with m even, a contradiction. Hence, the minimum value of $\|x\|$ for odd x_i is $\sqrt{32}$ and is achievable only by lattice points of shape $\pm 1^{23} \pm 3$.

Now suppose all the x_i are even. If $|x_i| > 4$ for some x_i, then $\Sigma\ x_i^2 \geq 6^2 > 32$, so we can disregard these vectors and assume $|x_i| = 0$, 2, or 4. If there is at least one x_i with $|x_i| = 2$, then there are at least eight (by examining the shapes of the codewords). Therefore $\Sigma\ x^2{}_i \geq 8 \cdot 2^2 = 32$. If always $|x_i| = 0$ or 4, then $|x_i| = 4$ for at least two x_i, or else $\Sigma\ x_i = \pm 4 = 4(\pm 1)$, contradicting the fact that m is even. If $|x_i| = 4$ for more than two x_i, then $\|x\|$ is too large. Thus, the minimum value of $\|x\|$ for even x_i is $\sqrt{32}$ and is achievable only by lattice points of shapes $0^{16} \pm 2^8$ and $0^{22} \pm 4^2$.

Theorem. The distance of Leech's lattice is $\sqrt{32}$. The neighbors of the origin have shapes $\pm 1^{23} \pm 3$, $0^{16} \pm 2^8$, and $0^{22} \pm 4^2$.

We now calculate the contact number $c(\mathcal{L})$ of Leech's lattice, noting first that each lattice point comes from a unique choice of c and m. The codewords that generate these lattice points have weights 8 or 16. Recalling Table 6.3, there are 759 codewords of weight 8 and 759 of weight 16. Since the lattice points of shape $0^{16} \pm 2^8$ satisfy the m even condition, there are an even number of +2 and −2 components. Suppose that the number of +2′s is a and the number of −2′s is $8 - a$. Then the possibilities for (a, b) are (8, 0), (2, 6), (4, 4), (6, 2), (0, 8). The first, third, and fifth of these come from codewords of weight 8 and the others from codewords of weight 16. Thus, there are $2^7 \cdot 759$ lattice points of shape $0^{16} \pm 2^8$.

How many lattice points have shape $0^{22} \pm 4^2$? Any choice of signs for the two 4′s satisfies the m even condition. Therefore, because there are $\binom{24}{2}$ choices for the placement of the 4′s (lattice points of this shape come from the all-0 codeword and the all-1 codeword), there are $2^2\binom{24}{2}$ lattice points of shape $0^{22} \pm 4^2$.

To find the number of lattice points of shape $\pm 1^{23} \pm 3$, let z be the number of +1′s in a lattice vector. Then $4m = z(1) - (23 - z) \pm 3 = 2z - 23 \pm 3$. This equation forces the +3 to occur if z is even and the –3 if z is odd. Any of the 2^{12} codewords generates a choice of +1′s and –1′s. The position of the ±3 may be chosen in 24 ways, and once it is chosen the sign is determined by the previous comment. Thus, there are $2^{12} \cdot 24$ lattice points of shape $\pm 1^{23} \pm 3$.

Therefore $c(\mathcal{L}) = 2^7 \cdot 759 + 2^2\binom{24}{2} + 2^{12} \cdot 24 = 196{,}560$.

This is the highest possible contact number for a lattice packing in $\mathbf{R}^{24}$, as proved by A. M. Odlyzko and N. J. A. Sloane in 1979.

Theorem. The contact number of Leech's lattice is 196560. Neighbors of the origin occur with the following multiplicities:

shape	number
$0^{16} \pm 2^8$	$2^7 \cdot 759 = 97152$
$0^{22} \pm 4^2$	$2^2\binom{24}{2} = 1104$
$\pm 1^{23} \pm 3$	$2^{12} \cdot 24 = 98304.$

We end the discussion with this respectable achievement, although the story of combinatorial designs is far from finished. The interested reader can turn to [27] and [8] for a description of the groups which J. H. Conway found in connection with $\mathcal{L}$. Conway denoted the automorphism group of $\mathcal{L}$ by $\cdot 0$ (pronounced "dotto") and defined $\cdot 1$ to be $\cdot 0$ divided by its center. It turns out that $\cdot 1$, along with two other automorphism groups, $\cdot 2$ and $\cdot 3$, are sporadic simple groups. The order bf $\cdot 1$ is

$$2^{21} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23.$$

Most simple groups belong to well-known families such as A_n, PSL(n, q), or the groups of Lie type. However, 26 simple groups do not fall into these categories and are therefore called *sporadic simple groups*. Several of the 26 sporadic simple groups are associated with Aut $\mathcal{L}$, including the largest, the *Monster group*, a group of symmetries in a space of 196,884 dimensions. The Monster group has order

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71.$$

EXERCISES

6.35 What is the highest possible contact number in a lattice packing in R^4?

6.36 Show how to use the (8, 16, 4) extended Hamming code to construct a lattice packing in $\mathbf{R}^8$ with contact number 240.

The relationship between the (8, 16, 4) code and the lattice E_8, with highest contact number (240) in $\mathbf{R}^8$, is discussed in [8].

Notes

In 1987 Luc Teirlinck proved the existence of t-designs for all values of t. However, there are many open problems. See the survey "Simple t-Designs with large t," by Donald L. Kreher, at http://www.math.mtu.edu/~kreher/.

Fisher's inequality was proved by the statistician and biologist Ronald A. Fisher (1890–1962). The Bruck-Chowla Ryser theorem was proved in 1949–1950 by Richard H. Bruck (1914–1991), Sarvadaman D. S. Chowla (1907–1995), and Herbert J. Ryser (1923–1985).

Philip Hall (1904–1982) published the marriage theorem in 1935. See [13] for a description of some equivalent theorems, such as the König–Egerváry theorem and Menger's theorem. The König–Egerváry theorem states that the minimum number of rows and columns which cover the 1′s in a 0–1 matrix equals the maximum number of row- and column-independent 1′s in the matrix. Menger's theorem states that the minimum number of points which separate two given nonadjacent points in a finite connected graph equals the maximum number of edge-disjoint paths which connect the two

points.

The problem of finding the highest possible contact number for a lattice in $\mathbf{R}^k$ is very much unsolved. Conway and Sloane [8] call this the *kissing number problem* and give a wealth of results. They also consider the related *packing problem*, the *covering problem* (in which one wants the least dense covering), and the so-called *quantizing problem* (which has application to data compression). These subjects abound with open questions. Although the obvious packing in $\mathbf{R}^3$ was proved by R. Hoppe in 1874 to have the highest possible contact number (see [1]), it was not proved until recently to be the densest possible packing. This result, known as Kepler's conjecture, was proved in 1998 by Thomas Hales at the University of Michigan. See http://www.math.lsa.umich.edurhales/.

In $\mathbf{R}^4$, the lattice packing with highest contact number, 24, is the D_4 lattice. We may take one sphere with center at the origin and the other spheres with centers of the form (±1, ±1, 0, 0), where we can take any combination of signs and the 0′s can occur in any two of the four coordinates. This accounts for $2^2 \cdot \binom{4}{2} = 24$ neighbors of the origin. These neighbors are the vertices of a *24-cell*, one of the six 4-dimensional Platonic solids.

In $\mathbf{R}^8$, the lattice packing with highest contact number is the E_8 lattice. Let C be the (8, 16, 4) extended Hamming code. It has 14 codewords of weight 4. Define the lattice to be the set of 8-tuples $(x_1,\ldots,x_8)$ such that $x_i \equiv c_i \pmod 2$ for $1 \le i \le 8$ and each codeword $c_i \in C$. The neighbors of the origin have the form (±2, 0, 0, 0, 0, 0, 0, 0), where the ±2 can be in any of the eight positions, or (±1, ±1, ±1, ±1, 0, 0, 0, 0), where we can take any combination of signs and the four 0′s correspond to the positions of 1′s in the weight 4 code words. There are $2 \cdot 8 + 14 \cdot 2^4 = 240$ neighbors of the origin. The number of lattice points at distances 0, 2, 4, 6,… from the origin are the coefficients of the remarkable theta series formula

$$\theta(q) = 1 + 240\sum_{n=1}^{\infty} \sigma_3(n) q^{2n} = 1 + 240q^2 + 2160q^4 + 6720q^6 + \cdots,$$

where $\sigma_3(n)$ is the sum of the cubes of the positive divisors of n.

In 1861 and 1873 Émile Mathieu (1835–1890) discovered five sporadic simple groups, M_{11}, M_{12}, M_{22}, M_{23}, and M_{24}, which are related to Steiner systems. The groups M_{24} and M_{23} are the automorphism groups of the $S(5, 8, 24)$ and $S(4, 7, 23)$ Steiner systems, respectively. The group M_{22} has index 2 in the automorphism group of the $S(3, 6, 22)$ Steiner system. The groups M_{12} and M_{11} are the automorphism groups of the $S(5, 6, 12)$ and $S(4, 5, 11)$ Steiner systems, respectively. These groups may also be defined in terms of permutations. For example, letting x = (1 2 3 4 5 6 7 8 9 10 11), y = (5 6 4 10)(11 8 3 7), and z = (1 12)(2 11)(3 6)(4 8)(5 9)(7 10), we can write $M_{11} = \langle x, y\rangle$ and $M_{12} = \langle x, y, z\rangle$. We have $|M_{11}| = 8 \cdot 9 \cdot 10 \cdot 11$ and $|M_{12}| = 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12$. With 7920 elements, M_{11} is the smallest of the 26 sporadic simple groups.

Mathieu group	order
M_{11}	$2^4 \cdot 3^2 \cdot 5 \cdot 11$
M_{12}	$2^6 \cdot 3^3 \cdot 5 \cdot 11$
M_{22}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$
M_{23}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23$
M_{24}	$2^{10} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 23$

APPENDIX A

WEB RESOURCES

Websites:

- http://www.combinatorics.org/Surveys/

  "The Electronic Journal of Combinatorics: Dynamic Surveys," a site containing up-to-date studies of combinatorics. The article *Small Ramsey Numbers*, by Stanislaw Radziszowski, shows the known values and bounds for Ramsey numbers.

- http://en.wikipedia.org/wiki/Combinatorics

  "Wikipedia: Combinatorics," a site developed by users, containing material about various aspects of combinatorics and links to other sources.

- http://en.wikipedia.org/wiki/Discrete_mathematics

  "Wikipedia: Discrete Mathematics," a site developed by users, containing material about various aspects of discrete mathematics and links to other sources.

- http://mathworld.wolfram.com/Combinatorics.html

  "MathWorld: Discrete Mathematics," a site containing material about various aspects of combinatorics and links to other sources.

- http://mathworld.wolfram.com/DiscreteMathematics.html

  "MathWorld: Discrete Mathematics," a site containing material about various aspects of discrete mathematics and links to other sources.

- http://oeis.org/

  "The On-Line Encyclopedia of Integer Sequences," a site containing integer sequences contributed by users. The sequences may be searched by initial terms, by name, or by attributes.

APPENDIX B

NOTATION

$n!$	n factorial, p. 4
$\mathbf{N}$	{1, 2, 3,…}, p. 6
$\mathbf{N}^m$	{1, 2, 3,…,m}, p. 6
$\binom{n}{k}$	binomial coefficient, p. 8
$\binom{n}{k_1,k_2,\ldots,k_m}$	multinomial coefficient, p. 9
$B(n)$	Bell number, p. 22
$p(n)$	partition number, p. 22
$p(n, k)$	partition number, p. 22
d_n	derangement number, p. 24
F_n	Fibonacci number, p. 31
L_n	Lucas number, p. 37
$\left\{ {n \atop k} \right\}$	Stirling number of the second kind, p. 40
$\left[{n \atop k} \right]$	Stirling number of the first kind, p. 41
$d_b(n)$	digital sum, p. 46
C_n	Catalan number, p. 64
t_n	number of transitive and reflexive relations, p. 69
p_n	number of partial orders, p. 69
$\mathbf{Z}_n$	cyclic group, p. 86
S_n	symmetric group, p. 87
D_n	dihedral group, p. 88
A_n	alternating group, p. 88
$g(n)$	number of nonisomorphic graphs, p. 101
$n(k, d)$	lattice point function, p. 112
$n'(k, d)$	generalized SET® function, p. 112
$\delta(g)$	degree of vertex, p. 115
$\overline{G}$	complement graph, p. 115
K_n	complete graph, p. 115
$K_{m,n}$	complete bipartite graph, p. 115
K_∞	infinite complete graph, p. 115
$K_{\infty,\infty}$	infinite complete bipartite graph, p. 115
C_n	cycle, p. 115
P_n	path, p. 115
$\alpha(G)$	independence number, p. 116
$\chi(G)$	chromatic number, p. 116
$R(m, n)$	Ramsey number, p. 133

$R(a_1,\ldots,a_c)$	multiple color Ramsey number, p. 135
$R(a_1,\ldots,a_c; t)$	hypergraph Ramsey number, p. 136
$[S]^t$	complete t-uniform hypergraph, 136
$S(c)$	Schur number, p. 145
$W(c, l)$	van der Waerden number, p. 148
FC	Fano Configuration, p. 162
GL(n, q)	general linear group, p. 164
SL(n, q)	special linear group, p. 164
PGL(n, q)	projective general linear group, p. 165
PSL(n, q)	projective special linear group, p. 165
t-(v, k, λ)	t-design, p. 171
$S(t, k, v)$	Steiner system, p. 173
(v, b, r, k, λ) BIBD	balanced incomplete block design, p. 175
(v, k, λ) SBD	square block design, p. 177
π_n	projective plane, p. 180
π'_n	affine plane, p. 182
L	Latin square, p. 183
MOLS	mutually orthogonal Latin squares, p. 185
OOD	ordered orthogonal design, p. 188
G_{23}	Golay code, p. 194
G_{24}	extended Golay code, p. 194
$\mathcal{L}$	Leech's lattice, p. 200

EXERCISE SOLUTIONS

SOLUTIONS FOR CHAPTER 1

1.5 There are 2^{99} binary strings of length 99. Half of these, 2^{98}, have an odd number of 1′s. The reason is that, regardless of the first 98 bits, the last bit may be chosen to make the total number of 1′s even or odd.

1.7 There are n^n such functions.

1.17 There are 2^{n^2} different binary relations on an n-set.

1.20 Since the tenth row of Pascal's triangle is 1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1, we have $(a + b)^{10} = a^{10} + 10a^9b + 45a^8b^2 + 120a^7b^3+210a^6b^4 + 252a^5b^5 + 210a^4b^6 + 120a^3b^7 + 45a^2b^8 + 10ab^9 + b^{10}$.

1.21 By the binomial theorem, the coefficient is $\binom{20}{10} = 184{,}756$.

1.25 By the multinomial theorem, $(a + b + c)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4 + 4a^3c\ 12a^2bc + 12ab^2c + 4b^3c + 6a^2c^2 + 12abc^2 + 6b^2c^2 + 4ac^3 + 4bc^3 + c^4$.

1.26 By the multinomial theorem, the coefficient is $\binom{20}{3,7,10} = 22{,}170{,}720$.

1.27 By simplifying, we see that

$$\binom{n}{k_1, k_2, \ldots, k_m} = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{n-k_1-k_2-\cdots-k_{m-2}}{k_{m-1}}.$$

1.30 (a) The number of paths is $\binom{25}{10,15} = 3{,}268{,}760$.

(b) The number of paths is $\binom{45}{10,15,20} = 10{,}361{,}546{,}974{,}682{,}663{,}760$.

1.32 The identity to prove is equivalent to

$$\sum_{k=2}^{n+1}\binom{k}{2} = \binom{n+2}{3}.$$

The right side is the number of selections of three distinct elements from the set $\{1,\ldots,n + 2\}$. Let the largest such element be $k + 1$, where $2 \le k \le n + 1$. Then, for each k, the other two elements must be selected from the set $\{1,\ldots,k\}$, and the number of ways to do this is $\binom{k}{2}$. The left side counts these selections for all k.

1.45 Clearly, $S_0(n) = n$. The formula for $S_1(n)$ can be obtained by noting that $2S_1(n) = (n + 1)n$ and hence $S_1(n) = n(n + 1)/2$. To find $S_2(n)$ and $S_3(n)$, we use the following technique. The sum $\Sigma^n_{i=1}[(i + 1)^{k+1} - i^{k+1}]$ is a telescoping series. Hence

$$(n+1)^{k+1}-1 = \sum_{i=1}^{n}[(i+1)^{k+1}-i^{k+1}] = \sum_{i=1}^{n}\sum_{j=0}^{k}\binom{k+1}{j}i^j = \sum_{j=0}^{k}\binom{k+1}{j}S_j(n).$$

Therefore

$$S_k(n) = \frac{1}{k+1}\left[(n+1)^{k+1} - 1 - \sum_{j=0}^{k-1}\binom{k+1}{j}S_j(n)\right].$$

Now we find

$$S_2(n) = \frac{1}{3}\left[(n+1)^3 - 1 - \binom{3}{0}n - \binom{3}{1}\frac{n(n+1)}{2}\right] = \frac{n(n+1)(2n+1)}{6}$$

and

$$S_3(n) = \frac{1}{4}\left[(n+1)^4 - 1 - \binom{4}{0}n - \binom{4}{1}\frac{n(n+1)}{2} - \binom{4}{2}\frac{n(n+1)(2n+1)}{6}\right]$$
$$= \left(\frac{n(n+1)}{2}\right)^2.$$

The fact that $S_k(n)$ is a monic polynomial in n of degree $k+1$ is clear from our method.

1.46 A k-dimensional face of an n-dimensional hypercube is a subset of the 2^n vertices of the cube isomorphic to a k-dimensional hypercube. There are $\binom{n}{k}$ choices for the coordinates on which to base the face. The other coordinates can be 0 or 1.

1.47 The total number of pizzas is

$$1 + 12 + \binom{2+12-1}{2} + \binom{3+12-1}{3} + \binom{4+12-1}{4} = 1820.$$

1.49 (a) The number of such functions is $\binom{n}{m}$.

(b) The number of such functions is $\binom{m+n-1}{m}$.

1.50 We have $\left\{{5 \atop 1}\right\} = 1$, $= \left\{{5 \atop 2}\right\} = 15$, $\left\{{5 \atop 3}\right\} = 25$, $\left\{{5 \atop 4}\right\} = 10$, $\left\{{5 \atop 5}\right\} = 1$, and $B(5) = 52$.

1.52 The formula $\left\{{n \atop 2}\right\} = 2^{n-1} - 1$ holds since any subset of $\{1, 2, 3,\ldots,n-1\}$, except the set itself, together with n, constitutes a part in a partition of $\{1, 2, 3,\ldots,n\}$ into two subsets. The formula $\left\{{n \atop n-1}\right\} = \binom{n}{2}$ for $n \geq 2$ holds since in a partition of $\{1, 2, 3,\ldots,n\}$ into $n-1$ parts, the element n is in any one of $n-1$ parts and there are $\binom{n}{2}$ choices for which part it is in.

1.53 There are $10^{30}/4 + 1$ choices for z, namely, all the integers from 0 to $10^{30}/4$. Among these choices, the average value of $4z$ is $10^{30}/2$. Hence, on average, $x+2y = 10^{30}/2$. There are $10^{30}/4 + 1$ values of y that satisfy this equation, namely, the integers from 0 to $10^{30}/2$. Once z and y are chosen, the value of x is determined. Therefore, the total number of nonnegative integer solutions is

$$\left(\frac{10^{30}}{4} + 1\right)^2.$$

1.56 (a) Let S be the set of functions from $\{1, 2, 3,\ldots,m\}$ to $\{1, 2, 3,\ldots,n\}$. Then S has cardinality n^m. We wish to find the cardinality of the subset of S consisting of all onto functions. For $1 \leq i \leq n$, let A_i be the collection of functions whose range does not contain i. Then the intersection of j of the A_i has cardinality $(n-j)^m$, the number of unrestricted functions from $\{1, 2, 3,\ldots,m\}$ to the $n-j$ nonexcluded elements of $\{1, 2, 3,\ldots,n\}$. Applying the inclusion–exclusion principle, we obtain

$$|A_1 \cup \cdots \cup A_n| = \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} (n-j)^m.$$

The number of onto functions is the complement of this union:

$$n^m - |A_1 \cup \cdots \cup A_n| = n^m - \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} (n-j)^m = \sum_{j=0}^{n} (-1)^j \binom{n}{j} (n-j)^m.$$

(b) From (a), we obtain

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n, \quad 1 \le k \le n.$$

1.60 We take successive differences:

1, 4, 13, 34, 73, 136, ...
3, 9, 21, 39, 63, ...
6, 12, 18, 24, ...
6, 6, 6, 6, ...

Having obtained a constant sequence, we stop. We find that the polynomial is

$$p(n) = 1 + 3n + 6\binom{n}{2} + 6\binom{n}{3} = n^3 + 2n + 1.$$

1.61 We can think of a pair of linked sets (*A, B*) as an onto function from *S* to the set {*A, B, A* ∩ *B*} or to the set $\{A, B, A \cap B, \overline{A \cup B}\}$. So the total number of pairs of linked sets is *T*(*n*, 3) + *T*(*n*, 4).

1.65 The Fibonacci numbers are sums of "shallow triangles" of Pascal's triangle:

$$\binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots = F_{n+1}, \quad n \ge 0.$$

We prove the identity by induction. For $n = 0$, we have

$$\binom{0}{0} = 1 = F_1.$$

Assume that the identity holds for n and $n - 1$. Then

$$\begin{aligned}
&\binom{n+1}{0} + \binom{n}{1} + \binom{n-1}{2} + \cdots \\
&= \binom{n}{0} + \left[\binom{n-1}{0} + \binom{n-1}{1}\right] + \left[\binom{n-2}{1} + \binom{n-2}{2}\right] + \cdots \\
&= \left[\binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots\right] \\
&\quad + \left[\binom{n-1}{0} + \binom{n-2}{1} + \binom{n-3}{2} + \cdots\right] \\
&= F_{n+1} + F_n \\
&= F_{n+2}.
\end{aligned}$$

Hence, the identity holds for $n + 1$ and by induction for all n.

1.68 The number 210 occurs six times in Pascal's triangle:

$$210 = \binom{210}{1} = \binom{210}{209} = \binom{10}{3} = \binom{10}{7} = \binom{16}{2} = \binom{16}{14}.$$

1.74 A particular solution to the recurrence relation is $a_n = -n - 3$. Hence, the general solution is of the form

$$a_n = A\phi^n + B\hat{\phi}^n - n - 3.$$

We need to choose A and B so that the initial values are satisfied. Thus

$$0 = A + B - 3$$

$$1 = A\phi + B\hat{\phi} - 4.$$

We find that $A = (15 + 7\sqrt{5})/10$ and $B = (15 - 7\sqrt{5})/10$.

1.86 We have $\left[{n \atop n-1}\right] = \binom{n}{2}$, as this counts the ways of choosing two elements to be in the same cycle.

1.88 In a partition of $n + 1$ elements, the $(n + 1)$st element is together with $n - k$ other elements for some k with $0 \le k \le n$. There are $\binom{n}{n-k}$ choices for these $n - k$ elements. Hence

$$B(n+1) = \sum_{k=0}^{n} \binom{n}{n-k} B(k) = \sum_{k=0}^{n} \binom{n}{k} B(k), \quad n \ge 0.$$

1.89 From the recurrence relation for Stirling numbers of the second kind,

$$k\left\{{n \atop k}\right\} = \left\{{n+1 \atop k}\right\} - \left\{{n \atop k-1}\right\},$$

and hence

$$\begin{aligned}\sum_{k=1}^{n} k\left\{{n \atop k}\right\} &= \sum_{k=1}^{n} \left(\left\{{n+1 \atop k}\right\} - \left\{{n \atop k-1}\right\}\right) \\ &= \sum_{k=1}^{n} \left\{{n+1 \atop k}\right\} - \sum_{k=1}^{n} \left\{{n \atop k-1}\right\} \\ &= B(n+1) - 1 - (B(n) - 1).\end{aligned}$$

Therefore, the expected number of parts is

$$\frac{1}{B(n)} \sum_{k=1}^{n} k\left\{{n \atop k}\right\} = \frac{B(n+1) - B(n)}{B(n)}.$$

1.90 Upon the substitutions $k \to -n + 1$ and $n \to -k + 1$, the recurrence relation (1.32) becomes the recurrence relation (1.33).

1.91 The relation follows by subtracting 1 from each part in the partition of n into k parts.

1.107 Obviously, $F_m | F_m$. Applying the identity

$$F_{m+n} = F_m F_{n+1} + F_{m-1} F_n, \quad m \ge 1, n \ge 0$$

(Exercise 1.64 with $n = m$), we obtain

$$F_{2m} = F_{m+1} F_m + F_m F_{m-1},$$

and hence $F_m | F_{2m}$. Similarly, with $n = 2m$, we obtain

$$F_{3m} = F_{m+2m} = F_{m+1} F_{2m} + F_m F_{2m-1},$$

and hence $F_m | F_{3m}$. Continuing in this manner, we find that $F_m \mid F_{km}$ for all $k \ge 1$.

Now suppose that $n = mk + r$, with $0 \le r < m$. Then, applying the identity again, we obtain

$$F_n = F_{mk+r} = F_{mk+1} F_r + F_{mk} F_{r-1}.$$

If $F_m \mid F_n$, then, since $F_m \mid F_{mk}$, it follows that $F_m \mid F_{mk-+1}\, F_r$. But consecutive Fibonacci numbers are coprime, and so F_m and F_{mk+1} are coprime. Hence $F_m \mid Fr$, but this is possible only if $r = 0$ (since $F_m > F_r$). Therefore $m \mid n$.

1.108 Since $\gcd(a, b) \mid a$, it follows from the previous exercise that $F_{\gcd(a,b)} \mid F_a$. Likewise, $F_{\gcd(a,b)} \mid F_b$. Hence $F_{\gcd(a,b)} \mid \gcd(F_a, F_b)$.

There exist integers x and y such that $\gcd(a, b) = ax + by$. Without loss of generality, assume that x is negative and y is positive (they cannot both be positive). Hence

$$\gcd(a, b) + a(-x) = by.$$

By the identity of the previous exercise,

$$F_{by} = F_{\gcd(a,b)+a(-x)} = F_{\gcd(a,b)+1}F_{a(-x)} + F_{\gcd(a,b)}F_{a(-x)-1}.$$

Thus, $F_{\gcd(a,b)}F_{a(-x)+1}$ is a linear combination of F_a and F_b and hence divisible by $\gcd(F_a, F_b)$. Since F_a and $F_{a(-x)+1}$ are relatively prime, it follows that $\gcd(F_a, F_b) \mid F_{\gcd(a,b)}$.

Therefore $F_{\gcd(a,b)} = \gcd(F_a, F_b)$.

SOLUTIONS FOR CHAPTER 2

2.1 The sum is

$$(xf'(x))|_{x=\frac{1}{3}} = \frac{6}{5}.$$

2.4 The generating function is $x(x((1-x)^{-1})')' = x(x+1)(1-x)^{-3}$.

2.12 A recurrence formula is $a_0 = 1$, $a_1 = -3$, and $a_n = -3a_{n-1} - a_{n-2}$ for $n \geq 2$. Notice that $a_n = (-1)^n F_{2n+2}$.

2.14 The identity follows upon summing $f(x)$, with x replaced by $\omega^j x$ and weighted by ω^{-jp} for $0 \leq j \leq q-1$, using the relation $1 + \omega + \omega^2 + \cdots + \omega^{q-1} = 0$.

2.17 We can find a simple formula for $t(n)$ where n has any given remainder upon division by 12. Thus

$$\begin{aligned}
t(12k) &= 3k^2 \\
t(12k+1) &= 3k^2 + 2k \\
t(12k+2) &= 3k^2 + k \\
t(12k+3) &= 3k^2 + 3k + 1 \\
t(12k+4) &= 3k^2 + 2k \\
t(12k+5) &= 3k^2 + 4k + 1 \\
t(12k+6) &= 3k^2 + 3k + 1 \\
t(12k+7) &= 3k^2 + 5k + 2 \\
t(12k+8) &= 3k^2 + 4k + 1 \\
t(12k+9) &= 3k^2 + 6k + 3 \\
t(12k+10) &= 3k^2 + 5k + 2 \\
t(12k+11) &= 3k^2 + 7k + 4.
\end{aligned}$$

Since these formula are all quadratic polynomials, $t(n)$ satisfies the recurrence relation

$$t(12k+r) = 3t(12(k-1)+r) - 3t(12(k-2)+r) + t(12(k-3)+r).$$

The desired recurrence relation follows immediately.

2.19 First, we prove that the period of $\{t(n) \bmod m\}$ is at most $12m$. If n is even, we may write $n = 12m + 2r$ and we have

$$t(12m+2r) = \{(12m+2r)^2/48\} = 2m^2 + mr + \{(2r)^2/48\} \equiv t(2r).$$

Here, $\{x\}$ denotes the nearest integer to x. The odd case is similar; it follows that $t(12m+u) \equiv t(u) \bmod m$ for all u. Hence, the period of $\{t(n) \bmod m\}$ is a divisor of $12m$ (and therefore at most $12m$).

Second, we show that the period of $\{t(n) \bmod m\}$ is at most $12m$.

Suppose that the period is p, and let $k = p$ or $p - 1$ so that k is even. We then have $\{k^2/48\} = 0 \pmod m$ and $\{(k+2)^2/48\} = 0 \pmod m$ since $t(-1) = t(0) = t(1) = t(2) = 0$. Thus m divides $\{(k+2)^2/48\} - \{k^2/48\}$ which is nonzero because $p > 12$. This difference is less than $(k+2)^2/48 k^2/48 + 1$, which implies that $12m < k + 13 \le p + 13$, thus completing the proof since $12m$ is a multiple of p and $p > 12$.

2.21 We will prove the result by induction on n. The claim is true for $n = 0$ since $C_0 = 1$ and $0 = 2^1 - 1$. Assume that the claim holds for all Catalan numbers up to C_n. Now consider C_{n+1}. If $n + 1$ is even, then

$$C_{n+1} = \sum_{k=0}^{(n-1)/2} C_k C_{n-k},$$

which is even. If $n + 1$ is odd, then

$$C_{n+1} = \sum_{k=0}^{n/2} C_k C_{n-k} + C_{n/2}^2.$$

So C_{m+1} is odd if and only if $C_{n/2}$ is odd. By the induction hypothesis, this means that $n/2 = 2^k - 1$, for some integer k. Thus, C_n is odd if and only if

$$n + 1 = 2(2^k - 1) + 1 = 2^{k+1} - 1.$$

By mathematical induction, the claim holds for all nonnegative integers n.

2.22 From the recurrence relation $C_n = \frac{4n-2}{n+1} C_{n-1}$, we have

$$(n+1)C_n \equiv (n+1)C_{n-1} \pmod 3.$$

If $n \not\equiv 2 \pmod 3$, then $C_n \not\equiv C_{n-1} \pmod 3$. Letting $n = 3k$, we have $C_{3k} \not\equiv C_{3k-1} \pmod 3$. Letting $n = 3k + 1$, we have $C_{3k+1} \not\equiv C_{3k} \pmod 3$. Therefore

$$C_{3k-1} \equiv C_{3k} \equiv C_{3k+1} \pmod 3.$$

2.27 Clearly, the formula holds for $n = 0$ and $n = 1$. Assume that it holds for n. Then

$$
\begin{aligned}
(x+y)^{(n+1)} &= (x+y)^{(n)}[(x+y)+n] \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{(k)} y^{(n-k)} [(x+k)+y+(n-k)] \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{(k+1)} y^{(n-k)} + \sum_{k=0}^{n} \binom{n}{k} x^{(k)} y^{(n-k+1)} \\
&= \sum_{k=1}^{n+1} \binom{n}{k-1} x^{(k)} y^{(n-k+1)} + \sum_{k=0}^{n} \binom{n}{k} x^{(k)} y^{(n-k+1)} \\
&= \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] x^{(k)} y^{(n+1-k)} \\
&\quad + y^{(n+1)} x^{(0)} + x^{(n+1)} y^{(0)} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} x^{(k)} y^{(n+1-k)}.
\end{aligned}
$$

Thus, the formula holds for $n + 1$ and hence for all n by induction.

The proof of the formula for $(x + y)_{(n)}$ is similar.

2.31 There are 48,639 such walks. It is easy to generate an 8 × 8 table by starting with 1 at a1 and at each cell adding the left, lower, and lower-left neighbors. This produces the Delannoy numbers. The main diagonal Delannoy numbers are 1, 3, 13, 63, 321, 1683, 8989, 48639.

2.32 A recurrence relation is

$$
\begin{aligned}
q(m,n) =& 2q(m-1,n) + 2q(m,n-1) - q(m-1,n-1) \\
& - 3q(m-2,n-1) - 3q(m-1,n-2) + 4q(m-2,n-2), \\
& m \geq 3 \text{ or } n \geq 3.
\end{aligned}
$$

The initial values are

$$
\begin{aligned}
&q(0,0) = 1; q(0,1) = 1; q(0,2) = 2 \\
&q(1,0) = 1; q(1,1) = 3; q(1,2) = 7 \\
&q(2,0) = 2; q(2,1) = 7; q(2,2) = 22.
\end{aligned}
$$

The corresponding generating function is

$$
\frac{1 - x - y + x^2 + xy^2 - x^2y^2}{1 - 2x - 2y + xy + 3x^2y + 3xy^2 - 4x^2y^2}.
$$

2.43 For $x \in X$, let N_x (neighborhood) be the intersection of all sets in S that contain x. Since S is closed under unions and complements, it is closed under intersections. Hence $N_x \in S$. The sets N_x partition X. Therefore, the number of algebras is equal to the number of partitions of X. (a) If X is labeled, then the number is $B(n)$. (b) If X is unlabeled, then the number is $p(n)$.

2.44 The number of functions is n^n. Given any element $x \in \{1,\ldots,n\}$, there are $(n–1)^n$ functions that do not include x in their range. Altogether, there are $n(n–1)^n$ of these elements x (with multiplicity). Hence

$$
r(n) = n - n\left(1 - \frac{1}{n}\right)^n,
$$

and so

$$\lim_{n\to\infty} \frac{r(n)}{n} = 1 - \frac{1}{e}.$$

2.52 There are 3210 such necklaces.

2.56 From the generating function in Example 2.22, we see that there are 16 circular necklaces with five white beads and five black beads.

2.63 There are 34 nonisomorphic graphs of order 5.

2.69 (a) We will show that the coefficients of $x_1^{\alpha_1} \cdots x_m^{\alpha_m}$ in both expressions are equal. Let $k = 1\alpha_1 + \cdots + m\alpha_m$ and $\alpha_{m+1} = \cdots = \alpha_k = 0$. In the expression on the left, the contribution to $x_1^{\alpha_1} \cdots x_m^{\alpha_m}$ comes from the $n = k$ term, which is

$$\frac{1}{k!} \frac{k!}{\prod_{j=1}^{k} j^{\alpha_j} \alpha_j!} = \frac{1}{\prod_{j=1}^{k} j^{\alpha_j} \alpha_j!}.$$

In the expression on the right, the factor $x_j^{\alpha_j}$ comes from the $x_j^{\alpha_j}/(j^{\alpha_j}\alpha_j!)$ term in the jth factor of $\exp \sum_{i=1}^{\infty}(x_i/i) = \prod_{i=1}^{\infty} \sum_{h=0}^{\infty} x_i^h/(i^h h!)$, for $j = 1, \cdots m$. Hence, the overall contribution is

$$\frac{1}{\prod_{j=1}^{k} j^{\alpha_j} \alpha_j!}.$$

2.70 There are two such graphs of order 5 and none of order 6. A self-complementary graph of order $n > 1$ has $n(n-1)/4$ edges. For this to be an integer, n must be of the form $4k$ or $k + 1$.

2.74 It is approximately 1.3×10^{1332}.

SOLUTIONS FOR CHAPTER 3

3.1 By the pigeonhole principle, there exist m and n, with $m < n$, such that $17^m \not\equiv 17^n \pmod{10^4}$. Since $\gcd(17, 10^4) = 1$, we have $17^{n-m} \not\equiv 1 \pmod{10^4}$.

We can actually find such an exponent using Euler's theorem: $a^{\phi(m)} \not\equiv 1 \pmod{m}$ if $\gcd(a, m) = 1$. The furnished exponent is $\phi(10^4) = 4000$.

3.3 To show that this result is the best possible, let $S = \{1,\ldots,100\}$ and take A_1 to be any subset of S with 67 elements. Let the other A_i be cyclic shifts of A_1. In this system of sets, each $x \in S$ is contained in exactly 67 of the A_i.

3.23 Let G_1 consist of three disjoint copies of K_4 and G_2 be G with one edge added.

3.24 If $\alpha(G)$ and $\chi(G)$ were both finite, then the number of vertices in G would be bounded by their product (a finite number). Since the number of vertices is infinite, at least one of $\alpha(G)$ and $\chi(G)$ is infinite.

3.25 Given any two vertices x and y of G, the degree conditions imply that x and y have a common neighbor z, and hence there is a path from x to y. Therefore G is connected.

3.26 Suppose that the longest path contained in G had fewer than n vertices. Then, choosing any two consecutive vertices in the path, say, x and y, the degree condition implies that x and y have a common neighbor z. Using z, we can make a longer path. Hence, the longest path in G has n vertices. The same type of argument allows us to close the path to make a cycle of length n.

3.27 The result follows by mathematical induction. Removing an end vertex and its incident edge from the graph amounts to subtracting 1 from both sides of the relation $p = q+1$. Continue this

process until the graph contains only one vertex; the relation certainly holds for this graph.

3.33 Take n disjoint copies of a total order on m elements.

SOLUTIONS FOR CHAPTER 4

4.2 We can take $f(n) = R(n, n)$. Number the vertices $1, 2, \ldots, f(n)$, and color the edge $\{i, j\}$, where $i < j$, green if i is directed to j, and red otherwise. Ramsey's theorem guarantees a monochromatic subgraph K_n. The corresponding directed subgraph is a transitive subtournament.

4.10 Take $G = K_{\infty, \infty, \infty}$. Call the parts of the tripartite graph A, B, and C. Color all edges between A and B green and all other edges red.

4.11 Let the graph be the cycle C_8 together with all four diagonals.

4.13 Given a 3-coloring of K_{17}, let v be any vertex. There are 16 edges incident with v. By the pigeonhole principle, at least six of these edges are the same color, say, green. Suppose that v is incident to vertices $x_1, x_2, x_3, x_4, x_5, x_6$ by green edges. If any edge $\{x_i, x_j\}$ is green, then we have a green K_3. If not, then we have a red K_3 or a blue K_3 (since $R(3, 3) = 6$). To finish the argument, show a 3-coloring of K_{16} with no monochromatic triangle.

4.15 From the probabilistic method, we obtain the lower bound 3.0038×10^{16}.

4.23 False. Take one part to be the union of $\{1\}$, $\{3, 4\}$, $\{7, 8, 9\}$, $\{13, 14, 15, 16\}$, etc.

4.26 An example of a 6-AP of primes is 7, 37, 67, 97, 127, 157.

SOLUTIONS FOR CHAPTER 5

5.1 The contribution from each component to both sides of the identity is 0 if $x_i = y_i$ and 1 if $x_i \neq y_i$.

5.4 0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111 The code detects one error.

5.6 Such a code would require that

$$19 \cdot \left[1 + \binom{10}{1} + \binom{10}{2}\right] \leq 2^{10},$$

a contradiction.

5.11 The probability that the Hamming code corrects one error is

$$\alpha = (1 - q)^7 + 7(1 - q)^6 q \doteq 0.96.$$

If no code is used, then the probability that the message is transmitted correctly is

$$\beta = (1 - q)^4 \doteq 0.81.$$

5.12 Append an eighth component to the Hamming code such that the entry makes the total number of 1′s even. This ensures that the distance of the code is 4.

5.25 The group is GL(3, 2).

SOLUTIONS FOR CHAPTER 6

6.2 A projective place of order 4 is a 2-(21, 5, 1) design.

6.7 Let the first row of the matrix be [1 1 0 1 0 0 0] and take cyclic shifts.

6.11 If $|C_i| = \lambda$ for some i, then every other C_j contains C_i, so $m \leq n + 1 - \lambda \leq n$.

Suppose not. Let $\gamma_i = |C_i| - \lambda > 0$ for $1 \le i \le m$. For the incidence matrix A, we have

$$\begin{aligned}\det(AA^t) &= \begin{vmatrix} \lambda+\gamma_1 & \lambda & \lambda & \cdots & \lambda \\ \lambda & \lambda+\gamma_2 & \lambda & \cdots & \lambda \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda & & \lambda & \lambda+\gamma_{m-1} & \lambda \\ \lambda & \cdots & \lambda & \lambda & \lambda+\gamma_m \end{vmatrix} \\ &= \gamma_1 \ldots \gamma_m[1 + \lambda(1/\gamma_1 + \cdots + 1/\gamma_m)] \neq 0.\end{aligned}$$

6.20 The result follows from the observation that, for any given α and β modulo n, there exist unique (mod n) i and j such that $i + j \equiv \alpha$ and $i - j \equiv \beta$. For $2i \equiv \alpha + \beta$ and $2j \equiv \alpha - \beta$, this determines unique i and j since n is odd.

6.21

1	3	2	4		1	2	3	4		1	2	3	4
2	4	1	3		2	1	4	3		4	3	2	1
4	2	3	1		3	4	1	2		2	1	4	3
3	1	4	2		4	3	2	1		3	4	1	2

6.33 Each lattice point has all n integer coordinates. Its lattice neighbors are those points with the same coordinates save for one change to an integer one greater or one lesser. Hence, each lattice point has $2n$ neighbors.

REFERENCES

1. M. Aigner and G. M. Ziegler. *Proofs from THE BOOK*. Springer-Verlag, New York, third edition, 2004.

2. N. Alon, J. Spencer, and P. Erdős. *The Probabilistic Method*. Wiley, New York, 1992.

3. I. Anderson. *Combinatorics of Finite Sets*. Oxford University Press, New York, 1987.

4. I. Anderson. *A First Course in Combinatorial Mathematics*. Oxford University Press, New York, second edition, 1989.

5. B. Bollobás. *Graph Theory: An Introductory Course*. Cambridge University Press, New York, 1979.

6. F. Chung and R. L. Graham. *Erdős on Graphs: His Legacy of Unsolved Problems*. A. K. Peters, Wellesley, second edition, 1999.

7. J. H. Conway. *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups*. Oxford University Press, New York, 1986.

8. J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices, and Groups*. Springer-Verlag, New York, third edition, 1999.

9. I. Gessel and G.-C. Rota, editors. *Classic Papers in Combinatorics*. Birkhäuser, Boston, 1987.

10. R. L. Graham, M. Grötschel, and L. Lovász, editors. *Handbook of Combinatorics*, volume 2. MIT Press, New York, 1995.

11. R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey Theory*. Wiley, New York, second edition, 1990.

12. M. Hall. *Combinatorial Theory*. Wiley, New York, second edition, 1986.

13. F. Harary. *Graph Theory*. Addison-Wesley, Reading, MA, 1969.

14. F. Harary and E. M. Palmer. *Graphical Enumeration*. Academic Press, New York, 1973.

15. G. Hardy and E. Wright. *An Introduction to the Theory of Numbers*. Clarendon Press, Oxford, fifth edition, 1989.

16. N. Hartsfield and G. Ringel. *Pearls in Graph Theory: A Comprehensive Introduction*. Academic Press, New York, second edition, 1994.

17. D. L. Johnson. *Presentations of Groups*. Cambridge University Press, New York, 1990.

18. V. Klee and S. Wagon. *Old and New Unsolved Problems in Plane Geometry and Number Theory*. Mathematical Association of America, Ithaca, NY, 1991.

19. J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge University Press, Cambridge, 1992.

20. R. C. Lyndon. *Groups and Geometry*. Cambridge University Press, New York, reprinted with corrections edition, 1986.

21. I. Niven, H. Zuckerman, and H. Montgomery. *An Introduction to the Theory of Numbers*. Wiley, New York, fifth edition, 1991.

22. V. Pless. *Introduction to the Theory of Error-Correcting Codes.* Wiley, New York, third edition, 1998.

23. J. Rotman. *The Theory of Groups: An Introduction.* Allyn and Bacon, Boston, second edition, 1973.

24. N. J. A. Sloane and S. Plouffe. *The Encyclopedia of Integer Sequences.* Academic Press, San Diego, 1995.

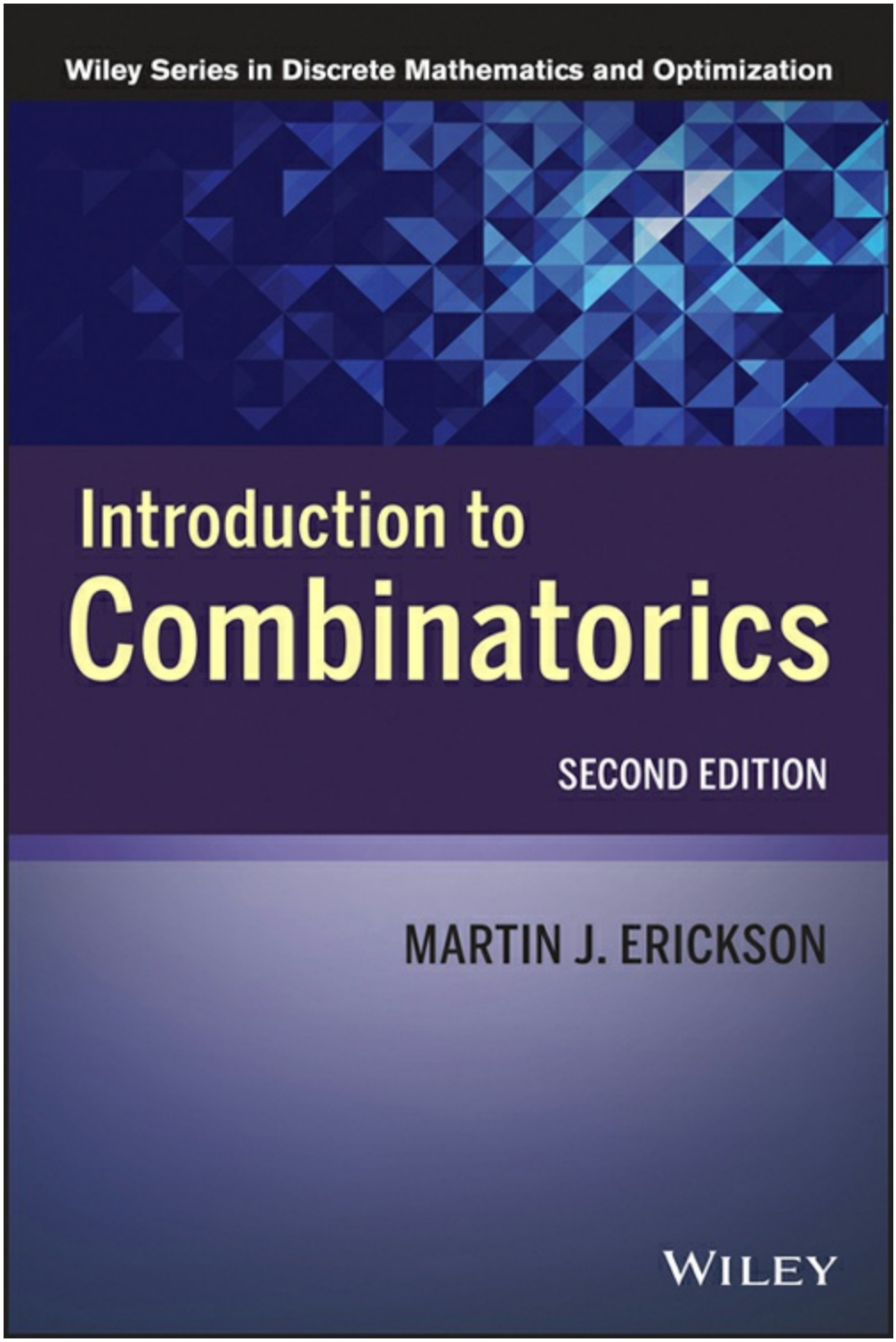
25. R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge University Press, New York, 1999.

26. D. W. Stanton, R. Stanton, and D. E. White. *Constructive Combinatorics.* Springer-Verlag, New York, 1986.

27. T. M. Thompson. *From Error-Correcting Codes Through Sphere Packings to Simple Groups.* Mathematical Association of America, Washington, DC, 1984.

28. D. B. West. *Introduction to Graph Theory.* Prentice Hall, Upper Saddle River, NJ, 1995.

Index


Wiley Series in Discrete Mathematics and Optimization
Introduction to
Combinatorics
SECOND EDITION
MARTIN J. ERICKSON
WILEY