CHAPTER

6 Counting

6.1 The Basics of Counting

6.2 The Pigeonhole Principle

6.3 Permutations and Combinations

6.4 Binomial Coefficients and Identities

6.5 Generalized Permutations and Combinations

6.6 Generating Permutations and Combinations

Combinatorics, the study of arrangements of objects, is an important part of discrete mathematics. This subject was studied as long ago as the seventeenth century, when combinatorial questions arose in the study of gambling games. Enumeration, the counting of objects with certain properties, is an important part of combinatorics. We must count objects to solve many different types of problems. For instance, counting is used to determine the complexity of algorithms. Counting is also required to determine whether there are enough telephone numbers or Internet protocol addresses to meet demand. Recently, it has played a key role in mathematical biology, especially in sequencing DNA. Furthermore, counting techniques are used extensively when probabilities of events are computed.

The basic rules of counting, which we will study in Section 6.1, can solve a tremendous variety of problems. For instance, we can use these rules to enumerate the different telephone numbers possible in the United States, the allowable passwords on a computer system, and the different orders in which the runners in a race can finish. Another important combinatorial tool is the pigeonhole principle, which we will study in Section 6.2. This states that when objects are placed in boxes and there are more objects than boxes, then there is a box containing at least two objects. For instance, we can use this principle to show that among a set of 15 or more students, at least 3 were born on the same day of the week.

We can phrase many counting problems in terms of ordered or unordered arrangements of the objects of a set with or without repetitions. These arrangements, called permutations and combinations, are used in many counting problems. For instance, suppose the 100 top finishers on a competitive exam taken by 2000 students are invited to a banquet. We can count the possible sets of 100 students that will be invited, as well as the ways in which the top 10 prizes can be awarded.

Another problem in combinatorics involves generating all the arrangements of a specified kind. This is often important in computer simulations. We will devise algorithms to generate arrangements of various types.

6.1 The Basics of Counting

6.1.1 Introduction

Suppose that a password on a computer system consists of six, seven, or eight characters. Each of these characters must be a digit or a letter of the alphabet. Each password must contain at least one digit. How many such passwords are there? The techniques needed to answer this question and a wide variety of other counting problems will be introduced in this section.

Counting problems arise throughout mathematics and computer science. For example, we must count the successful outcomes of experiments and all the possible outcomes of these experiments to determine probabilities of discrete events. We need to count the number of operations used by an algorithm to study its time complexity.

We will introduce the basic techniques of counting in this section. These methods serve as the foundation for almost all counting techniques.

6.1.2 Basic Counting Principles

Assessment

We first present two basic counting principles, the **product rule** and the **sum rule**. Then we will show how they can be used to solve many different counting problems.

The product rule applies when a procedure is made up of separate tasks.

> THE PRODUCT RULE Suppose that a procedure can be broken down into a sequence of two tasks. If there are n_1 ways to do the first task and for each of these ways of doing the first task, there are n_2 ways to do the second task, then there are $n_1 n_2$ ways to do the procedure.

Extra Examples

Examples 1–10 show how the product rule is used.

EXAMPLE 1 A new company with just two employees, Sanchez and Patel, rents a floor of a building with 12 offices. How many ways are there to assign different offices to these two employees?

Solution: The procedure of assigning offices to these two employees consists of assigning an office to Sanchez, which can be done in 12 ways, then assigning an office to Patel different from the office assigned to Sanchez, which can be done in 11 ways. By the product rule, there are $12 \cdot 11 = 132$ ways to assign offices to these two employees. ◀

EXAMPLE 2 The chairs of an auditorium are to be labeled with an uppercase English letter followed by a positive integer not exceeding 100. What is the largest number of chairs that can be labeled differently?

Solution: The procedure of labeling a chair consists of two tasks, namely, assigning to the seat one of the 26 uppercase English letters, and then assigning to it one of the 100 possible integers. The product rule shows that there are $26 \cdot 100 = 2600$ different ways that a chair can be labeled. Therefore, the largest number of chairs that can be labeled differently is 2600. ◀

EXAMPLE 3 There are 32 computers in a data center in the cloud. Each of these computers has 24 ports. How many different computer ports are there in this data center?

Solution: The procedure of choosing a port consists of two tasks, first picking a computer and then picking a port on this computer. Because there are 32 ways to choose the computer and 24 ways to choose the port no matter which computer has been selected, the product rule shows that there are $32 \cdot 24 = 768$ ports. ◀

An extended version of the product rule is often useful. Suppose that a procedure is carried out by performing the tasks $T_1, T_2, \ldots, T_m$ in sequence. If each task T_i, $i = 1, 2, \ldots, n$, can be done in n_i ways, regardless of how the previous tasks were done, then there are $n_1 \cdot n_2 \cdot \cdots \cdot n_m$ ways to carry out the procedure. This version of the product rule can be proved by mathematical induction from the product rule for two tasks (see Exercise 76).

EXAMPLE 4 How many different bit strings of length seven are there?

Solution: Each of the seven bits can be chosen in two ways, because each bit is either 0 or 1. Therefore, the product rule shows there are a total of $2^7 = 128$ different bit strings of length seven. ◀

EXAMPLE 5 How many different license plates can be made if each plate contains a sequence of three uppercase English letters followed by three digits (and no sequences of letters are prohibited, even if they are obscene)?

$$\underbrace{\_\ \_\ \_}_{\substack{\text{26 choices}\\ \text{for each}\\ \text{letter}}}\quad\underbrace{\_\ \_\ \_}_{\substack{\text{10 choices}\\ \text{for each}\\ \text{digit}}}$$

Solution: There are 26 choices for each of the three uppercase English letters and 10 choices for each of the three digits. Hence, by the product rule there are a total of $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17{,}576{,}000$ possible license plates. ◀

EXAMPLE 6 **Counting Functions** How many functions are there from a set with m elements to a set with n elements?

Solution: A function corresponds to a choice of one of the n elements in the codomain for each of the m elements in the domain. Hence, by the product rule there are $n \cdot n \cdot \cdots \cdot n = n^m$ functions from a set with m elements to one with n elements. For example, there are $5^3 = 125$ different functions from a set with three elements to a set with five elements. ◀

EXAMPLE 7 **Counting One-to-One Functions** How many one-to-one functions are there from a set with m elements to one with n elements?

Solution: First note that when $m > n$ there are no one-to-one functions from a set with m elements to a set with n elements.

Counting the number of onto functions is harder. We'll do this in Chapter 8.

Now let $m \le n$. Suppose the elements in the domain are $a_1, a_2, \ldots, a_m$. There are n ways to choose the value of the function at a_1. Because the function is one-to-one, the value of the function at a_2 can be picked in $n - 1$ ways (because the value used for a_1 cannot be used again). In general, the value of the function at a_k can be chosen in $n - k + 1$ ways. By the product rule, there are $n(n-1)(n-2)\cdots(n-m+1)$ one-to-one functions from a set with m elements to one with n elements.

For example, there are $5 \cdot 4 \cdot 3 = 60$ one-to-one functions from a set with three elements to a set with five elements. ◀

EXAMPLE 8 **The Telephone Numbering Plan** The *North American numbering plan (NANP)* specifies the format of telephone numbers in the U.S., Canada, and many other parts of North America. A telephone number in this plan consists of 10 digits, which are split into a three-digit area code, a three-digit office code, and a four-digit station code. Because of signaling considerations, there are certain restrictions on some of these digits. To specify the allowable format, let X denote a digit that can take any of the values 0 through 9, let N denote a digit that can take any of the values 2 through 9, and let Y denote a digit that must be a 0 or a 1. Two numbering plans, which will be called the old plan, and the new plan, will be discussed. (The old plan, in use in the 1960s, has been replaced by the new plan, but the recent rapid growth in demand for new numbers for mobile phones and devices will eventually make even this new plan obsolete. In this example, the letters used to represent digits follow the conventions of the *North American Numbering Plan*.) As will be shown, the new plan allows the use of more numbers.

Links ❯

Current projections are that by 2038, it will be necessary to add one or more digits to North American telephone numbers.

In the old plan, the formats of the area code, office code, and station code are *NYX*, *NNX*, and *XXXX*, respectively, so that telephone numbers had the form *NYX-NNX-XXXX*. In the new plan, the formats of these codes are *NXX*, *NXX*, and *XXXX*, respectively, so that telephone numbers have the form *NXX-NXX-XXXX*. How many different North American telephone numbers are possible under the old plan and under the new plan?

Solution: By the product rule, there are $8 \cdot 2 \cdot 10 = 160$ area codes with format *NYX* and $8 \cdot 10 \cdot 10 = 800$ area codes with format *NXX*. Similarly, by the product rule, there are

$8 \cdot 8 \cdot 10 = 640$ office codes with format *NNX*. The product rule also shows that there are $10 \cdot 10 \cdot 10 \cdot 10 = 10{,}000$ station codes with format *XXXX*.

Note that we have ignored restrictions that rule out N11 station codes for most area codes.

Consequently, applying the product rule again, it follows that under the old plan there are

$$160 \cdot 640 \cdot 10{,}000 = 1{,}024{,}000{,}000$$

different numbers available in North America. Under the new plan, there are

$$800 \cdot 800 \cdot 10{,}000 = 6{,}400{,}000{,}000$$

different numbers available. ◀

EXAMPLE 9 What is the value of k after the following code, where $n_1, n_2, \ldots, n_m$ are positive integers, has been executed?

```
k := 0
for i_1 := 1 to n_1
   for i_2 := 1 to n_2
         .
         .
         .
      for i_m := 1 to n_m
          k := k + 1
```

Solution: The initial value of k is zero. Each time the nested loop is traversed, 1 is added to k. Let T_i be the task of traversing the ith loop. Then the number of times the loop is traversed is the number of ways to do the tasks $T_1, T_2, \ldots, T_m$. The number of ways to carry out the task T_j, $j = 1, 2, \ldots, m$, is n_j, because the jth loop is traversed once for each integer i_j with $1 \le i_j \le n_j$. By the product rule, it follows that the nested loop is traversed $n_1 n_2 \cdots n_m$ times. Hence, the final value of k is $n_1 n_2 \cdots n_m$. ◀

EXAMPLE 10 **Counting Subsets of a Finite Set** Use the product rule to show that the number of different subsets of a finite set S is $2^{|S|}$.

Solution: Let S be a finite set. List the elements of S in arbitrary order. Recall from Section 2.2 that there is a one-to-one correspondence between subsets of S and bit strings of length $|S|$. Namely, a subset of S is associated with the bit string with a 1 in the ith position if the ith element in the list is in the subset, and a 0 in this position otherwise. By the product rule, there are $2^{|S|}$ bit strings of length $|S|$. Hence, $|P(S)| = 2^{|S|}$. (Recall that we used mathematical induction to prove this fact in Example 10 of Section 5.1.) ◀

The product rule is often phrased in terms of sets in this way: If $A_1, A_2, \ldots, A_m$ are finite sets, then the number of elements in the Cartesian product of these sets is the product of the number of elements in each set. To relate this to the product rule, note that the task of choosing an element in the Cartesian product $A_1 \times A_2 \times \cdots \times A_m$ is done by choosing an element in A_1, an element in A_2, …, and an element in A_m. By the product rule it follows that

$$|A_1 \times A_2 \times \cdots \times A_m| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_m|.$$

EXAMPLE 11 **DNA and Genomes** The hereditary information of a living organism is encoded using deoxyribonucleic acid (DNA), or in certain viruses, ribonucleic acid (RNA). DNA and RNA are extremely complex molecules, with different molecules interacting in a vast variety of ways to enable living process. For our purposes, we give only the briefest description of how DNA and RNA encode genetic information.

DNA molecules consist of two strands consisting of blocks known as nucleotides. Each nucleotide contains subcomponents called **bases**, each of which is adenine (A), cytosine (C), guanine (G), or thymine (T). The two strands of DNA are held together by hydrogen bonds connecting different bases, with A bonding only with T, and C bonding only with G. Unlike DNA, RNA is single stranded, with uracil (U) replacing thymine as a base. So, in DNA the possible base pairs are A-T and C-G, while in RNA they are A-U, and C-G. The DNA of a living creature consists of multiple pieces of DNA forming separate chromosomes. A **gene** is a segment of a DNA molecule that encodes a particular protein. The entirety of genetic information of an organism is called its **genome**.

Sequences of bases in DNA and RNA encode long chains of proteins called amino acids. There are 22 essential amino acids for human beings. We can quickly see that a sequence of at least three bases are needed to encode these 22 different amino acid. First note, that because there are four possibilities for each base in DNA, A, C, G, and T, by the product rule there are $4^2 = 16 < 22$ different sequences of two bases. However, there are $4^3 = 64$ different sequences of three bases, which provide enough different sequences to encode the 22 different amino acids (even after taking into account that several different sequences of three bases encode the same amino acid).

Soon it won't be that costly to have your own genetic code found.

The DNA of simple living creatures such as algae and bacteria have between 10^5 and 10^7 links, where each link is one of the four possible bases. More complex organisms, such as insects, birds, and mammals, have between 10^8 and 10^{10} links in their DNA. So, by the product rule, there are at least 4^{10^5} different sequences of bases in the DNA of simple organisms and at least 4^{10^8} different sequences of bases in the DNA of more complex organisms. These are both incredibly huge numbers, which helps explain why there is such tremendous variability among living organisms. In the past several decades techniques have been developed for determining the genome of different organisms. The first step is to locate each gene in the DNA of an organism. The next task, called **gene sequencing**, is the determination of the sequence of links on each gene. (The specific sequence of links on these genes depends on the particular individual representative of a species whose DNA is analyzed.) For example, the human genome includes approximately 23,000 genes, each with 1000 or more links. Gene sequencing techniques take advantage of many recently developed algorithms and are based on numerous new ideas in combinatorics. Many mathematicians and computer scientists work on problems involving genomes, taking part in the fast moving fields of bioinformatics and computational biology. ◀

We now introduce the sum rule.

> THE SUM RULE If a task can be done either in one of n_1 ways or in one of n_2 ways, where none of the set of n_1 ways is the same as any of the set of n_2 ways, then there are $n_1 + n_2$ ways to do the task.

Example 12 illustrates how the sum rule is used.

EXAMPLE 12 Suppose that either a member of the mathematics faculty or a student who is a mathematics major is chosen as a representative to a university committee. How many different choices are there for this representative if there are 37 members of the mathematics faculty and 83 mathematics majors and no one is both a faculty member and a student?

Solution: There are 37 ways to choose a member of the mathematics faculty and there are 83 ways to choose a student who is a mathematics major. Choosing a member of the mathematics faculty is never the same as choosing a student who is a mathematics major because no one is both a faculty member and a student. By the sum rule it follows that there are $37 + 83 = 120$ possible ways to pick this representative. ◀

We can extend the sum rule to more than two tasks. Suppose that a task can be done in one of n_1 ways, in one of n_2 ways, … , or in one of n_m ways, where none of the set of n_i ways of doing the task is the same as any of the set of n_j ways, for all pairs i and j with $1 \leq i < j \leq m$. Then the number of ways to do the task is $n_1 + n_2 + \cdots + n_m$. This extended version of the sum rule is often useful in counting problems, as Examples 13 and 14 show. This version of the sum rule can be proved using mathematical induction from the sum rule for two sets. (See Exercise 75.)

EXAMPLE 13 A student can choose a computer project from one of three lists. The three lists contain 23, 15, and 19 possible projects, respectively. No project is on more than one list. How many possible projects are there to choose from?

Solution: The student can choose a project by selecting a project from the first list, the second list, or the third list. Because no project is on more than one list, by the sum rule there are $23 + 15 + 19 = 57$ ways to choose a project. ◀

EXAMPLE 14 What is the value of k after the following code, where $n_1, n_2, \ldots, n_m$ are positive integers, has been executed?

```
k := 0
for i_1 := 1 to n_1
    k := k + 1
for i_2 := 1 to n_2
    k := k + 1
        .
        .
        .
for i_m := 1 to n_m
    k := k + 1
```

Solution: The initial value of k is zero. This block of code is made up of m different loops. Each time a loop is traversed, 1 is added to k. To determine the value of k after this code has been executed, we need to determine how many times we traverse a loop. Note that there are n_i ways to traverse the ith loop. Because we only traverse one loop at a time, the sum rule shows that the final value of k, which is the number of ways to traverse one of the m loops is $n_1 + n_2 + \cdots + n_m$. ◀

The sum rule can be phrased in terms of sets as: If $A_1, A_2, \ldots, A_m$ are pairwise disjoint finite sets, then the number of elements in the union of these sets is the sum of the numbers of elements in the sets. To relate this to our statement of the sum rule, note there are $|A_i|$ ways to choose an element from A_i for $i = 1, 2, \ldots, m$. Because the sets are pairwise disjoint, when we select an element from one of the sets A_i, we do not also select an element from a different set A_j. Consequently, by the sum rule, because we cannot select an element from two of these sets at the same time, the number of ways to choose an element from one of the sets, which is the number of elements in the union, is

$$|A_1 \cup A_2 \cup \cdots \cup A_m| = |A_1| + |A_2| + \cdots + |A_m| \text{ when } A_i \cap A_j = \quad \text{for all } i, j.$$

This equality applies only when the sets in question are pairwise disjoint. The situation is much more complicated when these sets have elements in common. That situation will be briefly discussed later in this section and discussed in more depth in Chapter 8.

6.1.3 More Complex Counting Problems

Many counting problems cannot be solved using just the sum rule or just the product rule. However, many complicated counting problems can be solved using both of these rules in combination. We begin by counting the number of variable names in the programming language BASIC. (In the exercises, we consider the number of variable names in JAVA.) Then we will count the number of valid passwords subject to a particular set of restrictions.

EXAMPLE 15 In a version of the computer language BASIC, the name of a variable is a string of one or two alphanumeric characters, where uppercase and lowercase letters are not distinguished. (An *alphanumeric* character is either one of the 26 English letters or one of the 10 digits.) Moreover, a variable name must begin with a letter and must be different from the five strings of two characters that are reserved for programming use. How many different variable names are there in this version of BASIC?

Extra Examples

Solution: Let V equal the number of different variable names in this version of BASIC. Let V_1 be the number of these that are one character long and V_2 be the number of these that are two characters long. Then by the sum rule, $V = V_1 + V_2$. Note that $V_1 = 26$, because a one-character variable name must be a letter. Furthermore, by the product rule there are $26 \cdot 36$ strings of length two that begin with a letter and end with an alphanumeric character. However, five of these are excluded, so $V_2 = 26 \cdot 36 - 5 = 931$. Hence, there are $V = V_1 + V_2 = 26 + 931 = 957$ different names for variables in this version of BASIC. ◀

EXAMPLE 16 Each user on a computer system has a password, which is six to eight characters long, where each character is an uppercase letter or a digit. Each password must contain at least one digit. How many possible passwords are there?

Solution: Let P be the total number of possible passwords, and let P_6, P_7, and P_8 denote the number of possible passwords of length 6, 7, and 8, respectively. By the sum rule, $P = P_6 + P_7 + P_8$. We will now find P_6, P_7, and P_8. Finding P_6 directly is difficult. To find P_6 it is easier to find the number of strings of uppercase letters and digits that are six characters long, including those with no digits, and subtract from this the number of strings with no digits. By the product rule, the number of strings of six characters is 36^6, and the number of strings with no digits is 26^6. Hence,

$$P_6 = 36^6 - 26^6 = 2{,}176{,}782{,}336 - 308{,}915{,}776 = 1{,}867{,}866{,}560.$$

Similarly, we have

$$P_7 = 36^7 - 26^7 = 78{,}364{,}164{,}096 - 8{,}031{,}810{,}176 = 70{,}332{,}353{,}920$$

and

$$\begin{aligned} P_8 &= 36^8 - 26^8 = 2{,}821{,}109{,}907{,}456 - 208{,}827{,}064{,}576 \\ &= 2{,}612{,}282{,}842{,}880. \end{aligned}$$

Consequently,

$$P = P_6 + P_7 + P_8 = 2{,}684{,}483{,}063{,}360.$$ ◀

Bit Number	0	1	2	3	4	8	16	24	31
Class A	0	netid				hostid			
Class B	1	0	netid				hostid		
Class C	1	1	0	netid				hostid	
Class D	1	1	1	0	Multicast Address				
Class E	1	1	1	1	0	Address			

FIGURE 1 **Internet addresses (IPv4).**

EXAMPLE 17

Links

Counting Internet Addresses In the Internet, which is made up of interconnected physical networks of computers, each computer (or more precisely, each network connection of a computer) is assigned an *Internet address*. In Version 4 of the Internet Protocol (IPv4), still in use today, an address is a string of 32 bits. It begins with a *network number* (*netid*). The netid is followed by a *host number* (*hostid*), which identifies a computer as a member of a particular network.

Three forms of addresses are used, with different numbers of bits used for netids and hostids. **Class A addresses**, used for the largest networks, consist of 0, followed by a 7-bit netid and a 24-bit hostid. **Class B addresses**, used for medium-sized networks, consist of 10, followed by a 14-bit netid and a 16-bit hostid. **Class C addresses**, used for the smallest networks, consist of 110, followed by a 21-bit netid and an 8-bit hostid. There are several restrictions on addresses because of special uses: 1111111 is not available as the netid of a Class A network, and the hostids consisting of all 0s and all 1s are not available for use in any network. A computer on the Internet has either a Class A, a Class B, or a Class C address. (Besides Class A, B, and C addresses, there are also Class D addresses, reserved for use in multicasting when multiple computers are addressed at a single time, consisting of 1110 followed by 28 bits, and Class E addresses, reserved for future use, consisting of 11110 followed by 27 bits. Neither Class D nor Class E addresses are assigned as the IPv4 address of a computer on the Internet.) Figure 1 illustrates IPv4 addressing. (Limitations on the number of Class A and Class B netids have made IPv4 addressing inadequate; IPv6, a new version of IP, uses 128-bit addresses to solve this problem.)

The lack of available IPv4 address has become a crisis!

How many different IPv4 addresses are available for computers on the Internet?

Solution: Let x be the number of available addresses for computers on the Internet, and let x_A, x_B, and x_C denote the number of Class A, Class B, and Class C addresses available, respectively. By the sum rule, $x = x_A + x_B + x_C$.

To find x_A, note that there are $2^7 - 1 = 127$ Class A netids, recalling that the netid 1111111 is unavailable. For each netid, there are $2^{24} - 2 = 16{,}777{,}214$ hostids, recalling that the hostids consisting of all 0s and all 1s are unavailable. Consequently, $x_A = 127 \cdot 16{,}777{,}214 = 2{,}130{,}706{,}178$.

To find x_B and x_C, note that there are $2^{14} = 16{,}384$ Class B netids and $2^{21} = 2{,}097{,}152$ Class C netids. For each Class B netid, there are $2^{16} - 2 = 65{,}534$ hostids, and for each Class C netid, there are $2^8 - 2 = 254$ hostids, recalling that in each network the hostids consisting of all 0s and all 1s are unavailable. Consequently, $x_B = 1{,}073{,}709{,}056$ and $x_C = 532{,}676{,}608$.

We conclude that the total number of IPv4 addresses available is $x = x_A + x_B + x_C = 2{,}130{,}706{,}178 + 1{,}073{,}709{,}056 + 532{,}676{,}608 = 3{,}737{,}091{,}842$. ◀

6.1.4 The Subtraction Rule (Inclusion–Exclusion for Two Sets)

Suppose that a task can be done in one of two ways, but some of the ways to do it are common to both ways. In this situation, we cannot use the sum rule to count the number of ways to do the task. If we add the number of ways to do the tasks in these two ways, we get an overcount of the total number of ways to do it, because the ways to do the task that are common to the two ways are counted twice.

Overcounting is perhaps the most common enumeration error.

To correctly count the number of ways to do the two tasks, we must subtract the number of ways that are counted twice. This leads us to an important counting rule.

THE SUBTRACTION RULE If a task can be done in either n_1 ways or n_2 ways, then the number of ways to do the task is $n_1 + n_2$ minus the number of ways to do the task that are common to the two different ways.

The subtraction rule is also known as the **principle of inclusion–exclusion**, especially when it is used to count the number of elements in the union of two sets. Suppose that A_1 and A_2 are sets. Then, there are $|A_1|$ ways to select an element from A_1 and $|A_2|$ ways to select an element from A_2. The number of ways to select an element from A_1 or from A_2, that is, the number of ways to select an element from their union, is the sum of the number of ways to select an element from A_1 and the number of ways to select an element from A_2, minus the number of ways to select an element that is in both A_1 and A_2. Because there are $|A_1 \cup A_2|$ ways to select an element in either A_1 or in A_2, and $|A_1 \cap A_2|$ ways to select an element common to both sets, we have

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

This is the formula given in Section 2.2 for the number of elements in the union of two sets.

Example 18 illustrates how we can solve counting problems using the subtraction principle.

EXAMPLE 18 How many bit strings of length eight either start with a 1 bit or end with the two bits 00?

Extra Examples

Solution: Figure 2 illustrates the three counting problems we need to solve before we can apply the principle of inclusion–exclusion. We can construct a bit string of length eight that either starts with a 1 bit or ends with the two bits 00, by constructing a bit string of length eight beginning with a 1 bit or by constructing a bit string of length eight that ends with the two bits 00. We can construct a bit string of length eight that begins with a 1 in $2^7 = 128$ ways. This follows by the product rule, because the first bit can be chosen in only one way and each of the other seven bits can be chosen in two ways. Similarly, we can construct a bit string of length eight ending with the two bits 00, in $2^6 = 64$ ways. This follows by the product rule, because each of the first six bits can be chosen in two ways and the last two bits can be chosen in only one way.

Some of the ways to construct a bit string of length eight starting with a 1 are the same as the ways to construct a bit string of length eight that ends with the two bits 00. There are $2^5 = 32$ ways to construct such a string. This follows by the product rule, because the first bit can be chosen in only one way, each of the second through the sixth bits can be chosen in two ways, and the last two bits can be chosen in one way. Consequently, the number of bit strings of length eight that begin with a 1 or end with a 00, which equals the number of ways to construct a bit string of length eight that begins with a 1 or that ends with 00, equals $128 + 64 - 32 = 160$. ◀

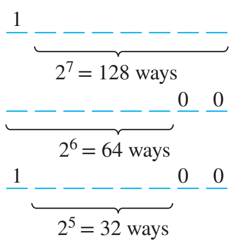


FIGURE 2 8-Bit strings starting with 1 or ending with 00.

We present an example that illustrates how the formulation of the principle of inclusion–exclusion can be used to solve counting problems.

EXAMPLE 19 A computer company receives 350 applications from college graduates for a job planning a line of new web servers. Suppose that 220 of these applicants majored in computer science, 147

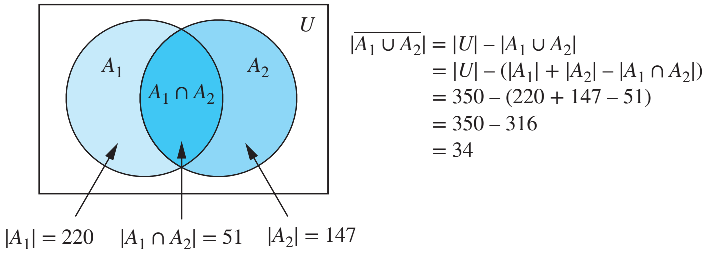


FIGURE 3 Applicants who majored in neither computer science nor business.

majored in business, and 51 majored both in computer science and in business. How many of these applicants majored neither in computer science nor in business?

Solution: To find the number of these applicants who majored neither in computer science nor in business, we can subtract the number of students who majored either in computer science or in business (or both) from the total number of applicants. Let A_1 be the set of students who majored in computer science and A_2 the set of students who majored in business. Then $A_1 \cup A_2$ is the set of students who majored in computer science or business (or both), and $A_1 \cap A_2$ is the set of students who majored both in computer science and in business. By the subtraction rule the number of students who majored either in computer science or in business (or both) equals

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| = 220 + 147 - 51 = 316.$$

We conclude that $350 - 316 = 34$ of the applicants majored neither in computer science nor in business. A Venn diagram for this example is shown in Figure 3. ◀

The subtraction rule, or the principle of inclusion–exclusion, can be generalized to find the number of ways to do one of n different tasks or, equivalently, to find the number of elements in the union of n sets, whenever n is a positive integer. We will study the inclusion–exclusion principle and some of its many applications in Chapter 8.

6.1.5 The Division Rule

We have introduced the product, sum, and subtraction rules for counting. You may wonder whether there is also a division rule for counting. In fact, there is such a rule, which can be useful when solving certain types of enumeration problems.

THE DIVISION RULE There are n/d ways to do a task if it can be done using a procedure that can be carried out in n ways, and for every way w, exactly d of the n ways correspond to way w.

We can restate the division rule in terms of sets: "If the finite set A is the union of n pairwise disjoint subsets each with d elements, then $n = |A|/d$."

We can also formulate the division rule in terms of functions: "If f is a function from A to B where A and B are finite sets, and that for every value $y \in B$ there are exactly d values $x \in A$ such that $f(x) = y$ (in which case, we say that f is d-to-one), then $|B| = |A|/d$."

Remark: The division rule comes in handy when it appears that a task can be done in n different ways, but it turns out that for each way of doing the task, there are d equivalent ways of doing it. Under these circumstances, we can conclude that there are n/d inequivalent ways of doing the task.

We illustrate the use of the division rule for counting with two examples.

EXAMPLE 20 Suppose that an automated system has been developed that counts the legs of cows in a pasture. Suppose that this system has determined that in a farmer's pasture there are exactly 572 legs. How many cows are there is this pasture, assuming that each cow has four legs and that there are no other animals present?

Solution: Let n be the number of cow legs counted in a pasture. Because each cow has four legs, by the division rule we know that the pasture contains $n/4$ cows. Consequently, the pasture with 572 cow legs has $572/4 = 143$ cows in it. ◀

EXAMPLE 21 How many different ways are there to seat four people around a circular table, where two seatings are considered the same when each person has the same left neighbor and the same right neighbor?

Solution: We arbitrarily select a seat at the table and label it seat 1. We number the rest of the seats in numerical order, proceeding clockwise around the table. Note that are four ways to select the person for seat 1, three ways to select the person for seat 2, two ways to select the person for seat 3, and one way to select the person for seat 4. Thus, there are $4! = 24$ ways to order the given four people for these seats. However, each of the four choices for seat 1 leads to the same arrangement, as we distinguish two arrangements only when one of the people has a different immediate left or immediate right neighbor. Because there are four ways to choose the person for seat 1, by the division rule there are $24/4 = 6$ different seating arrangements of four people around the circular table. ◀

6.1.6 Tree Diagrams

Counting problems can be solved using **tree diagrams**. A tree consists of a root, a number of branches leaving the root, and possible additional branches leaving the endpoints of other branches. (We will study trees in detail in Chapter 11.) To use trees in counting, we use a branch to represent each possible choice. We represent the possible outcomes by the leaves, which are the endpoints of branches not having other branches starting at them.

Note that when a tree diagram is used to solve a counting problem, the number of choices of which branch to follow to reach a leaf can vary as in Example 22.

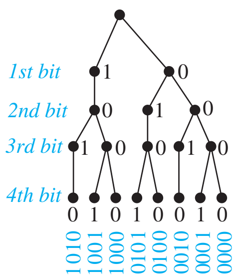


FIGURE 4 Bit strings of length four without consecutive 1s.

EXAMPLE 22 How many bit strings of length four do not have two consecutive 1s?

Solution: The tree diagram in Figure 4 displays all bit strings of length four without two consecutive 1s. We see that there are eight bit strings of length four without two consecutive 1s. ◀

EXAMPLE 23 A playoff between two teams consists of at most five games. The first team that wins three games wins the playoff. In how many different ways can the playoff occur?

Solution: The tree diagram in Figure 5 displays all the ways the playoff can proceed, with the winner of each game shown. We see that there are 20 different ways for the playoff to occur. ◀

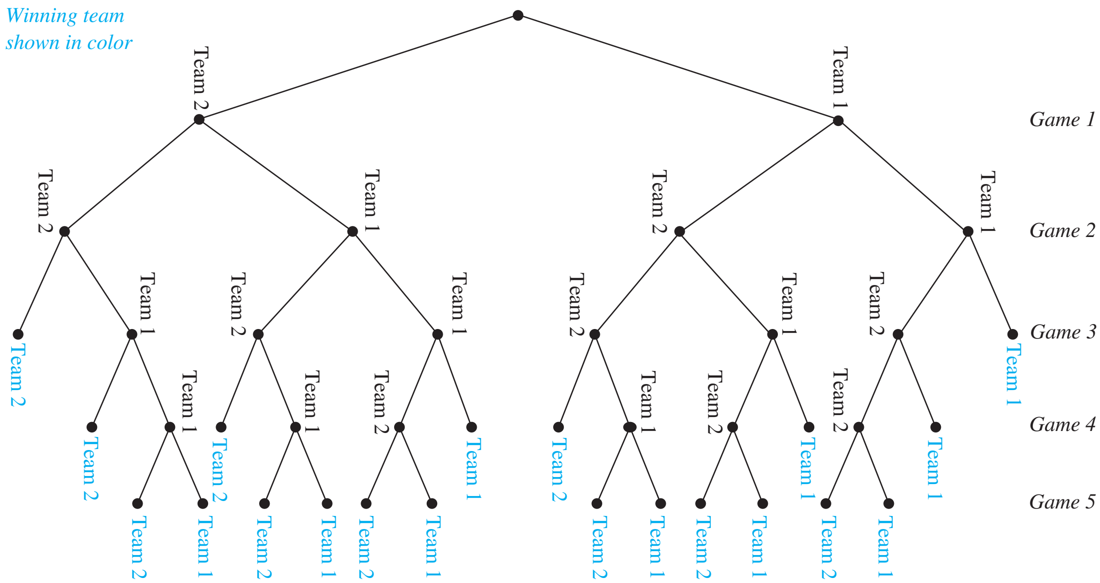


FIGURE 5 **Best three games out of five playoffs.**

EXAMPLE 24 Suppose that "I Love New Jersey" T-shirts come in five different sizes: S, M, L, XL, and XXL. Further suppose that each size comes in four colors, white, red, green, and black, except for XL, which comes only in red, green, and black, and XXL, which comes only in green and black. How many different shirts does a souvenir shop have to stock to have at least one of each available size and color of the T-shirt?

Solution: The tree diagram in Figure 6 displays all possible size and color pairs. It follows that the souvenir shop owner needs to stock 17 different T-shirts. ◀

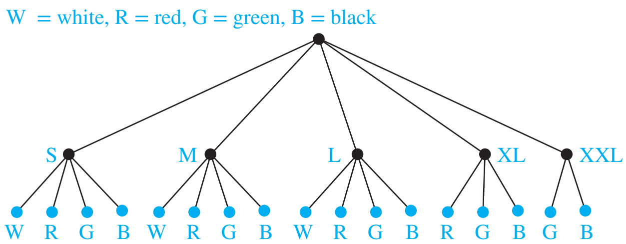


FIGURE 6 **Counting varieties of T-shirts.**

Exercises

1. There are 18 mathematics majors and 325 computer science majors at a college.

a) In how many ways can two representatives be picked so that one is a mathematics major and the other is a computer science major?

b) In how many ways can one representative be picked who is either a mathematics major or a computer science major?

2. An office building contains 27 floors and has 37 offices on each floor. How many offices are in the building?

3. A multiple-choice test contains 10 questions. There are four possible answers for each question.

a) In how many ways can a student answer the questions on the test if the student answers every question?

b) In how many ways can a student answer the questions on the test if the student can leave answers blank?

4. A particular brand of shirt comes in 12 colors, has a male version and a female version, and comes in three sizes for each sex. How many different types of this shirt are made?

5. Six different airlines fly from New York to Denver and seven fly from Denver to San Francisco. How many different pairs of airlines can you choose on which to book a trip from New York to San Francisco via Denver, when you pick an airline for the flight to Denver and an airline for the continuation flight to San Francisco?

6. There are four major auto routes from Boston to Detroit and six from Detroit to Los Angeles. How many major auto routes are there from Boston to Los Angeles via Detroit?

7. How many different three-letter initials can people have?

8. How many different three-letter initials with none of the letters repeated can people have?

9. How many different three-letter initials are there that begin with an A?

10. How many bit strings are there of length eight?

11. How many bit strings of length ten both begin and end with a 1?

12. How many bit strings are there of length six or less, not counting the empty string?

13. How many bit strings with length not exceeding n, where n is a positive integer, consist entirely of 1s, not counting the empty string?

14. How many bit strings of length n, where n is a positive integer, start and end with 1s?

15. How many strings are there of lowercase letters of length four or less, not counting the empty string?

16. How many strings are there of four lowercase letters that have the letter x in them?

17. How many strings of five ASCII characters contain the character @ ("at" sign) at least once? [*Note:* There are 128 different ASCII characters.]

18. How many 5-element DNA sequences

a) end with A?
b) start with T and end with G?
c) contain only A and T?
d) do not contain C?

19. How many 6-element RNA sequences

a) do not contain U?
b) end with GU?
c) start with C?
d) contain only A or U?

20. How many positive integers between 5 and 31

a) are divisible by 3? Which integers are these?
b) are divisible by 4? Which integers are these?
c) are divisible by 3 and by 4? Which integers are these?

21. How many positive integers between 50 and 100

a) are divisible by 7? Which integers are these?
b) are divisible by 11? Which integers are these?
c) are divisible by both 7 and 11? Which integers are these?

22. How many positive integers less than 1000

a) are divisible by 7?
b) are divisible by 7 but not by 11?
c) are divisible by both 7 and 11?
d) are divisible by either 7 or 11?
e) are divisible by exactly one of 7 and 11?
f) are divisible by neither 7 nor 11?
g) have distinct digits?
h) have distinct digits and are even?

23. How many positive integers between 100 and 999 inclusive

a) are divisible by 7?
b) are odd?
c) have the same three decimal digits?
d) are not divisible by 4?
e) are divisible by 3 or 4?
f) are not divisible by either 3 or 4?
g) are divisible by 3 but not by 4?
h) are divisible by 3 and 4?

24. How many positive integers between 1000 and 9999 inclusive

a) are divisible by 9?
b) are even?
c) have distinct digits?
d) are not divisible by 3?
e) are divisible by 5 or 7?
f) are not divisible by either 5 or 7?
g) are divisible by 5 but not by 7?
h) are divisible by 5 and 7?

25. How many strings of three decimal digits

a) do not contain the same digit three times?
b) begin with an odd digit?
c) have exactly two digits that are 4s?

26. How many strings of four decimal digits

a) do not contain the same digit twice?
b) end with an even digit?
c) have exactly three digits that are 9s?

27. A committee is formed consisting of one representative from each of the 50 states in the United States, where the representative from a state is either the governor or one of the two senators from that state. How many ways are there to form this committee?

28. How many license plates can be made using either three digits followed by three uppercase English letters or three uppercase English letters followed by three digits?

29. How many license plates can be made using either two uppercase English letters followed by four digits or two digits followed by four uppercase English letters?

30. How many license plates can be made using either three uppercase English letters followed by three digits or four uppercase English letters followed by two digits?

31. How many license plates can be made using either two or three uppercase English letters followed by either two or three digits?

32. How many strings of eight uppercase English letters are there

a) if letters can be repeated?
b) if no letter can be repeated?
c) that start with X, if letters can be repeated?
d) that start with X, if no letter can be repeated?
e) that start and end with X, if letters can be repeated?
f) that start with the letters BO (in that order), if letters can be repeated?
g) that start and end with the letters BO (in that order), if letters can be repeated?
h) that start or end with the letters BO (in that order), if letters can be repeated?

33. How many strings of eight English letters are there

a) that contain no vowels, if letters can be repeated?
b) that contain no vowels, if letters cannot be repeated?
c) that start with a vowel, if letters can be repeated?
d) that start with a vowel, if letters cannot be repeated?
e) that contain at least one vowel, if letters can be repeated?
f) that contain exactly one vowel, if letters can be repeated?
g) that start with X and contain at least one vowel, if letters can be repeated?
h) that start and end with X and contain at least one vowel, if letters can be repeated?

34. How many different functions are there from a set with 10 elements to sets with the following numbers of elements?

a) 2 **b)** 3 **c)** 4 **d)** 5

35. How many one-to-one functions are there from a set with five elements to sets with the following number of elements?

a) 4 **b)** 5 **c)** 6 **d)** 7

36. How many functions are there from the set $\{1, 2, \ldots, n\}$, where n is a positive integer, to the set $\{0, 1\}$?

37. How many functions are there from the set $\{1, 2, \ldots, n\}$, where n is a positive integer, to the set $\{0, 1\}$

a) that are one-to-one?
b) that assign 0 to both 1 and n?
c) that assign 1 to exactly one of the positive integers less than n?

38. How many partial functions (see Section 2.3) are there from a set with five elements to sets with each of these number of elements?

a) 1 **b)** 2 **c)** 5 **d)** 9

39. How many partial functions (see Definition 13 of Section 2.3) are there from a set with m elements to a set with n elements, where m and n are positive integers?

40. How many subsets of a set with 100 elements have more than one element?

41. A **palindrome** is a string whose reversal is identical to the string. How many bit strings of length n are palindromes?

42. How many 4-element DNA sequences

a) do not contain the base T?
b) contain the sequence ACG?
c) contain all four bases A, T, C, and G?
d) contain exactly three of the four bases A, T, C, and G?

43. How many 4-element RNA sequences

a) contain the base U?
b) do not contain the sequence CUG?
c) do not contain all four bases A, U, C, and G?
d) contain exactly two of the four bases A, U, C, and G?

44. On each of the 22 work days in a particular month, every employee of a start-up venture was sent a company communication. If a total of 4642 total company communications were sent, how many employees does the company have, assuming that no staffing changes were made that month?

45. At a large university, 434 freshmen, 883 sophomores, and 43 juniors are enrolled in an introductory algorithms course. How many sections of this course need to be scheduled to accommodate all these students if each section contains 34 students?

46. How many ways are there to seat four of a group of ten people around a circular table where two seatings are considered the same when everyone has the same immediate left and immediate right neighbor?

47. How many ways are there to seat six people around a circular table where two seatings are considered the same when everyone has the same two neighbors without regard to whether they are right or left neighbors?

48. In how many ways can a photographer at a wedding arrange 6 people in a row from a group of 10 people, where the bride and the groom are among these 10 people, if

a) the bride must be in the picture?
b) both the bride and groom must be in the picture?
c) exactly one of the bride and the groom is in the picture?

49. In how many ways can a photographer at a wedding arrange six people in a row, including the bride and groom, if

a) the bride must be next to the groom?
b) the bride is not next to the groom?
c) the bride is positioned somewhere to the left of the groom?

50. How many bit strings of length seven either begin with two 0s or end with three 1s?

51. How many bit strings of length 10 either begin with three 0s or end with two 0s?

***52.** How many bit strings of length 10 contain either five consecutive 0s or five consecutive 1s?

****53.** How many bit strings of length eight contain either three consecutive 0s or four consecutive 1s?

54. Every student in a discrete mathematics class is either a computer science or a mathematics major or is a joint major in these two subjects. How many students are in the class if there are 38 computer science majors (including joint majors), 23 mathematics majors (including joint majors), and 7 joint majors?

55. How many positive integers not exceeding 100 are divisible either by 4 or by 6?

56. How many different initials can someone have if a person has at least two, but no more than five, different initials? Assume that each initial is one of the 26 uppercase letters of the English language.

57. Suppose that a password for a computer system must have at least 8, but no more than 12, characters, where each character in the password is a lowercase English letter, an uppercase English letter, a digit, or one of the six special characters $*$, $>$, $<$, !, $+$, and $=$.

a) How many different passwords are available for this computer system?

b) How many of these passwords contain at least one occurrence of at least one of the six special characters?

c) Using your answer to part (a), determine how long it takes a hacker to try every possible password, assuming that it takes one nanosecond for a hacker to check each possible password.

58. The name of a variable in the C programming language is a string that can contain uppercase letters, lowercase letters, digits, or underscores. Further, the first character in the string must be a letter, either uppercase or lowercase, or an underscore. If the name of a variable is determined by its first eight characters, how many different variables can be named in C? (Note that the name of a variable may contain fewer than eight characters.)

59. The name of a variable in the JAVA programming language is a string of between 1 and 65,535 characters, inclusive, where each character can be an uppercase or a lowercase letter, a dollar sign, an underscore, or a digit, except that the first character must not be a digit. Determine the number of different variable names in JAVA.

60. The International Telecommunications Union (ITU) specifies that a telephone number must consist of a country code with between 1 and 3 digits, except that the code 0 is not available for use as a country code, followed by a number with at most 15 digits. How many available possible telephone numbers are there that satisfy these restrictions?

61. Suppose that at some future time every telephone in the world is assigned a number that contains a country code 1 to 3 digits long, that is, of the form X, XX, or XXX, followed by a 10-digit telephone number of the form NXX-NXX-$XXXX$ (as described in Example 8). How many different telephone numbers would be available worldwide under this numbering plan?

62. A key in the Vigenère cryptosystem is a string of English letters, where the case of the letters does not matter. How many different keys for this cryptosystem are there with three, four, five, or six letters?

63. A wired equivalent privacy (WEP) key for a wireless fidelity (WiFi) network is a string of either 10, 26, or 58 hexadecimal digits. How many different WEP keys are there?

64. Suppose that p and q are prime numbers and that $n = pq$. Use the principle of inclusion–exclusion to find the number of positive integers not exceeding n that are relatively prime to n.

65. Use the principle of inclusion–exclusion to find the number of positive integers less than 1,000,000 that are not divisible by either 4 or by 6.

66. Use a tree diagram to find the number of bit strings of length four with no three consecutive 0s.

67. How many ways are there to arrange the letters a, b, c, and d such that a is not followed immediately by b?

68. Use a tree diagram to find the number of ways that the World Series can occur, where the first team that wins four games out of seven wins the series.

69. Use a tree diagram to determine the number of subsets of $\{3, 7, 9, 11, 24\}$ with the property that the sum of the elements in the subset is less than 28.

70. **a)** Suppose that a store sells six varieties of soft drinks: cola, ginger ale, orange, root beer, lemonade, and cream soda. Use a tree diagram to determine the number of different types of bottles the store must stock to have all varieties available in all size bottles if all varieties are available in 12-ounce bottles, all but lemonade are available in 20-ounce bottles, only cola and ginger ale are available in 32-ounce bottles, and all but lemonade and cream soda are available in 64-ounce bottles?

b) Answer the question in part (a) using counting rules.

71. **a)** Suppose that a popular style of running shoe is available for both men and women. The woman's shoe comes in sizes 6, 7, 8, and 9, and the man's shoe comes in sizes 8, 9, 10, 11, and 12. The man's shoe comes in white and black, while the woman's shoe comes in white, red, and black. Use a tree diagram to determine the number of different shoes that a store has to stock to have at least one pair of this type of running shoe for all available sizes and colors for both men and women.

b) Answer the question in part (a) using counting rules.

72. Determine the number of matches played in a single-elimination tournament with n players, where for each game between two players the winner goes on, but the loser is eliminated.

73. Determine the minimum and the maximum number of matches that can be played in a double-elimination tournament with n players, where after each game between two players, the winner goes on and the loser goes on if and only if this is not a second loss.

***74.** Use the product rule to show that there are 2^{2^n} different truth tables for propositions in n variables.

75. Use mathematical induction to prove the sum rule for m tasks from the sum rule for two tasks.

76. Use mathematical induction to prove the product rule for m tasks from the product rule for two tasks.

77. How many diagonals does a convex polygon with n sides have? (Recall that a polygon is convex if every line segment connecting two points in the interior or boundary of the polygon lies entirely within this set and that a diagonal of a polygon is a line segment connecting two vertices that are not adjacent.)

78. Data are transmitted over the Internet in **datagrams**, which are structured blocks of bits. Each datagram contains header information organized into a maximum of 14 different fields (specifying many things, including the source and destination addresses) and a data area that contains the actual data that are transmitted. One of the 14 header fields is the **header length field** (denoted by HLEN), which is specified by the protocol to be 4 bits long and that specifies the header length in terms of 32-bit blocks of bits. For example, if HLEN $= 0110$, the header is made up of six 32-bit blocks. Another of the 14 header fields is the 16-bit-long **total length field** (denoted by TOTAL LENGTH), which specifies the length in bits of the entire datagram, including both the header fields and the data area. The length of the data area is the total length of the datagram minus the length of the header.

a) The largest possible value of TOTAL LENGTH (which is 16 bits long) determines the maximum total length in octets (blocks of 8 bits) of an Internet datagram. What is this value?

b) The largest possible value of HLEN (which is 4 bits long) determines the maximum total header length in 32-bit blocks. What is this value? What is the maximum total header length in octets?

c) The minimum (and most common) header length is 20 octets. What is the maximum total length in octets of the data area of an Internet datagram?

d) How many different strings of octets in the data area can be transmitted if the header length is 20 octets and the total length is as long as possible?

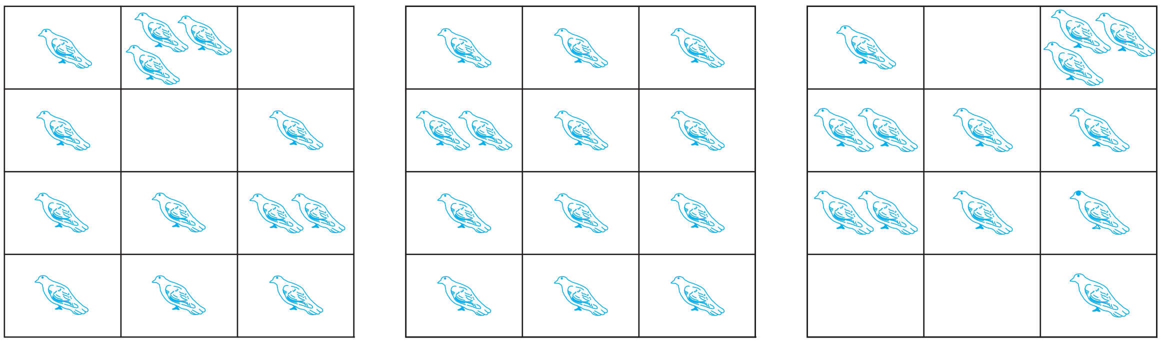
6.2 The Pigeonhole Principle

6.2.1 Introduction

Links

Suppose that a flock of 20 pigeons flies into a set of 19 pigeonholes to roost. Because there are 20 pigeons but only 19 pigeonholes, a least one of these 19 pigeonholes must have at least two pigeons in it. To see why this is true, note that if each pigeonhole had at most one pigeon in it, at most 19 pigeons, one per hole, could be accommodated. This illustrates a general principle called the **pigeonhole principle**, which states that if there are more pigeons than pigeonholes, then there must be at least one pigeonhole with at least two pigeons in it (see Figure 1). This principle is extremely useful; it applies to much more than pigeons and pigeonholes.

THEOREM 1 **THE PIGEONHOLE PRINCIPLE** If k is a positive integer and $k + 1$ or more objects are placed into k boxes, then there is at least one box containing two or more of the objects.



(a) (b) (c)

FIGURE 1 **There are more pigeons than pigeonholes.**

Proof: We prove the pigeonhole principle using a proof by contraposition. Suppose that none of the k boxes contains more than one object. Then the total number of objects would be at most k. This is a contradiction, because there are at least $k+1$ objects. ◁

The pigeonhole principle is also called the **Dirichlet drawer principle**, after the nineteenth-century German mathematician G. Lejeune Dirichlet, who often used this principle in his work. (Dirichlet was not the first person to use this principle; a demonstration that there were at least two Parisians with the same number of hairs on their heads dates back to the 17th century—see Exercise 35.) It is an important additional proof technique supplementing those we have developed in earlier chapters. We introduce it in this chapter because of its many important applications to combinatorics.

We will illustrate the usefulness of the pigeonhole principle. We first show that it can be used to prove a useful corollary about functions.

COROLLARY 1 A function f from a set with $k+1$ or more elements to a set with k elements is not one-to-one.

Proof: Suppose that for each element y in the codomain of f we have a box that contains all elements x of the domain of f such that $f(x)=y$. Because the domain contains $k+1$ or more elements and the codomain contains only k elements, the pigeonhole principle tells us that one of these boxes contains two or more elements x of the domain. This means that f cannot be one-to-one. ◁

Examples 1–3 show how the pigeonhole principle is used.

EXAMPLE 1 Among any group of 367 people, there must be at least two with the same birthday, because there are only 366 possible birthdays. ◀

EXAMPLE 2 In any group of 27 English words, there must be at least two that begin with the same letter, because there are 26 letters in the English alphabet. ◀

EXAMPLE 3 How many students must be in a class to guarantee that at least two students receive the same score on the final exam, if the exam is graded on a scale from 0 to 100 points?

Solution: There are 101 possible scores on the final. The pigeonhole principle shows that among any 102 students there must be at least 2 students with the same score. ◀

Links

G. LEJEUNE DIRICHLET (1805–1859) G. Lejeune Dirichlet was born into a Belgian family living near Cologne, Germany. His father was a postmaster. He became passionate about mathematics at a young age. He was spending all his spare money on mathematics books by the time he entered secondary school in Bonn at the age of 12. At 14 he entered the Jesuit College in Cologne, and at 16 he began his studies at the University of Paris. In 1825 he returned to Germany and was appointed to a position at the University of Breslau. In 1828 he moved to the University of Berlin. In 1855 he was chosen to succeed Gauss at the University of Göttingen. Dirichlet is said to be the first person to master Gauss's *Disquisitiones Arithmeticae*, which appeared 20 years earlier. He is said to have kept a copy at his side even when he traveled. Dirichlet made many important discoveries in number theory, including the theorem that there are infinitely many primes in arithmetical progressions $an+b$ when a and b are relatively prime. He proved the $n=5$ case of Fermat's last theorem, that there are no nontrivial solutions in integers to $x^5+y^5=z^5$. Dirichlet also made many contributions to analysis. Dirichlet was considered to be an excellent teacher who could explain ideas with great clarity. He was married to Rebecka Mendelssohn, one of the sisters of the composer Felix Mendelssohn.

The pigeonhole principle is a useful tool in many proofs, including proofs of surprising results, such as that given in Example 4.

EXAMPLE 4 Show that for every integer n there is a multiple of n that has only 0s and 1s in its decimal expansion.

Extra Examples

Solution: Let n be a positive integer. Consider the $n+1$ integers $1, 11, 111, \ldots, 11\ldots1$ (where the last integer in this list is the integer with $n+1$ 1s in its decimal expansion). Note that there are n possible remainders when an integer is divided by n. Because there are $n+1$ integers in this list, by the pigeonhole principle there must be two with the same remainder when divided by n. The larger of these integers less the smaller one is a multiple of n, which has a decimal expansion consisting entirely of 0s and 1s. ◀

6.2.2 The Generalized Pigeonhole Principle

The pigeonhole principle states that there must be at least two objects in the same box when there are more objects than boxes. However, even more can be said when the number of objects exceeds a multiple of the number of boxes. For instance, among any set of 21 decimal digits there must be 3 that are the same. This follows because when 21 objects are distributed into 10 boxes, one box must have more than 2 objects.

THEOREM 2 **THE GENERALIZED PIGEONHOLE PRINCIPLE** If N objects are placed into k boxes, then there is at least one box containing at least $\lceil N/k \rceil$ objects.

Proof: We will use a proof by contraposition. Suppose that none of the boxes contains more than $\lceil N/k \rceil - 1$ objects. Then, the total number of objects is at most

$$k\left(\left\lceil \frac{N}{k} \right\rceil - 1\right) < k\left(\left(\frac{N}{k} + 1\right) - 1\right) = N,$$

where the inequality $\lceil N/k \rceil < (N/k) + 1$ has been used. Thus, the total number of objects is less than N. This completes the proof by contraposition. ◁

A common type of problem asks for the minimum number of objects such that at least r of these objects must be in one of k boxes when these objects are distributed among the boxes. When we have N objects, the generalized pigeonhole principle tells us there must be at least r objects in one of the boxes as long as $\lceil N/k \rceil \geq r$. The smallest integer N with $N/k > r - 1$, namely, $N = k(r-1) + 1$, is the smallest integer satisfying the inequality $\lceil N/k \rceil \geq r$. Could a smaller value of N suffice? The answer is no, because if we had $k(r-1)$ objects, we could put $r-1$ of them in each of the k boxes and no box would have at least r objects.

When thinking about problems of this type, it is useful to consider how you can avoid having at least r objects in one of the boxes as you add successive objects. To avoid adding a rth object to any box, you eventually end up with $r-1$ objects in each box. There is no way to add the next object without putting an rth object in that box.

Examples 5–8 illustrate how the generalized pigeonhole principle is applied.

EXAMPLE 5 Among 100 people there are at least $\lceil 100/12 \rceil = 9$ who were born in the same month. ◀

EXAMPLE 6 What is the minimum number of students required in a discrete mathematics class to be sure that at least six will receive the same grade, if there are five possible grades, A, B, C, D, and F?

Extra Examples

Solution: The minimum number of students needed to ensure that at least six students receive the same grade is the smallest integer N such that $\lceil N/5 \rceil = 6$. The smallest such integer is $N = 5 \cdot 5 + 1 = 26$. If you have only 25 students, it is possible for there to be five who have received each grade so that no six students have received the same grade. Thus, 26 is the minimum number of students needed to ensure that at least six students will receive the same grade. ◂

EXAMPLE 7 **a)** How many cards must be selected from a standard deck of 52 cards to guarantee that at least three cards of the same suit are selected?
b) How many must be selected from a standard deck of 52 cards to guarantee that at least three hearts are selected?

A standard deck of 52 cards has 13 kinds of cards, with four cards of each of kind, one in each of the four suits, hearts, diamonds, spades, and clubs.

Solution: **a)** Suppose there are four boxes, one for each suit, and as cards are selected they are placed in the box reserved for cards of that suit. Using the generalized pigeonhole principle, we see that if N cards are selected, there is at least one box containing at least $\lceil N/4 \rceil$ cards. Consequently, we know that at least three cards of one suit are selected if $\lceil N/4 \rceil \geq 3$. The smallest integer N such that $\lceil N/4 \rceil \geq 3$ is $N = 2 \cdot 4 + 1 = 9$, so nine cards suffice. Note that if eight cards are selected, it is possible to have two cards of each suit, so more than eight cards are needed. Consequently, nine cards must be selected to guarantee that at least three cards of one suit are chosen. One good way to think about this is to note that after the eighth card is chosen, there is no way to avoid having a third card of some suit.

b) We do not use the generalized pigeonhole principle to answer this question, because we want to make sure that there are three hearts, not just three cards of one suit. Note that in the worst case, we can select all the clubs, diamonds, and spades, 39 cards in all, before we select a single heart. The next three cards will be all hearts, so we may need to select 42 cards to get three hearts. ◂

EXAMPLE 8 What is the least number of area codes needed to guarantee that the 25 million phones in a state can be assigned distinct 10-digit telephone numbers? (Assume that telephone numbers are of the form *NXX-NXX-XXXX*, where the first three digits form the area code, N represents a digit from 2 to 9 inclusive, and X represents any digit.)

Solution: There are eight million different phone numbers of the form *NXX-XXXX* (as shown in Example 8 of Section 6.1). Hence, by the generalized pigeonhole principle, among 25 million telephones, at least $\lceil 25{,}000{,}000/8{,}000{,}000 \rceil = 4$ of them must have identical phone numbers. Hence, at least four area codes are required to ensure that all 10-digit numbers are different. ◂

Example 9, although not an application of the generalized pigeonhole principle, makes use of similar principles.

EXAMPLE 9 Suppose that a computer science laboratory has 15 workstations and 10 servers. A cable can be used to directly connect a workstation to a server. For each server, only one direct connection to that server can be active at any time. We want to guarantee that at any time any set of 10 or fewer workstations can simultaneously access different servers via direct connections. Although we could do this by connecting every workstation directly to every server (using 150 connections), what is the minimum number of direct connections needed to achieve this goal?

Solution: Suppose that we label the workstations $W_1, W_2, \ldots, W_{15}$ and the servers $S_1, S_2, \ldots, S_{10}$. First, we would like to find a way for there to be far fewer than 150 direct connections between workstations and servers to achieve our goal. One promising approach is to directly connect W_k to S_k for $k = 1, 2, \ldots, 10$ and then to connect each of $W_{11}, W_{12}, W_{13}, W_{14}$, and W_{15} to all

10 servers. This gives us a total of $10 + 5 \cdot 10 = 60$ direct connections. We need to determine whether with this configuration any set of 10 or fewer workstations can simultaneously access different servers. We note that if workstation W_j is included with $1 \leq j \leq 10$, it can access server S_j, and for each workstation W_k with $k \geq 11$ included, there must be a corresponding workstation W_j with $1 \leq j \leq 10$ not included, so W_k can access server S_j. (This follows because there are at least as many available servers S_j as there are workstations W_j with $1 \leq j \leq 10$ not included.) So, any set of 10 or fewer workstations are able to simultaneously access different servers.

But can we use fewer than 60 direct connections? Suppose there are fewer than 60 direct connections between workstations and servers. Then some server would be connected to at most $\lfloor 59/10 \rfloor = 5$ workstations. (If all servers were connected to at least six workstations, there would be at least $6 \cdot 10 = 60$ direct connections.) This means that the remaining nine servers are not enough for the other 10 or more workstations to simultaneously access different servers. Consequently, at least 60 direct connections are needed. It follows that 60 is the answer. ◀

6.2.3 Some Elegant Applications of the Pigeonhole Principle

In many interesting applications of the pigeonhole principle, the objects to be placed in boxes must be chosen in a clever way. A few such applications will be described here.

EXAMPLE 10 During a month with 30 days, a baseball team plays at least one game a day, but no more than 45 games. Show that there must be a period of some number of consecutive days during which the team must play exactly 14 games.

Solution: Let a_j be the number of games played on or before the jth day of the month. Then $a_1, a_2, \ldots, a_{30}$ is an increasing sequence of distinct positive integers, with $1 \leq a_j \leq 45$. Moreover, $a_1 + 14, a_2 + 14, \ldots, a_{30} + 14$ is also an increasing sequence of distinct positive integers, with $15 \leq a_j + 14 \leq 59$.

The 60 positive integers $a_1, a_2, \ldots, a_{30}, a_1 + 14, a_2 + 14, \ldots, a_{30} + 14$ are all less than or equal to 59. Hence, by the pigeonhole principle two of these integers are equal. Because the integers $a_j, j = 1, 2, \ldots, 30$ are all distinct and the integers $a_j + 14, j = 1, 2, \ldots, 30$ are all distinct, there must be indices i and j with $a_i = a_j + 14$. This means that exactly 14 games were played from day $j + 1$ to day i. ◀

EXAMPLE 11 Show that among any $n + 1$ positive integers not exceeding $2n$ there must be an integer that divides one of the other integers.

Solution: Write each of the $n + 1$ integers $a_1, a_2, \ldots, a_{n+1}$ as a power of 2 times an odd integer. In other words, let $a_j = 2^{k_j} q_j$ for $j = 1, 2, \ldots, n + 1$, where k_j is a nonnegative integer and q_j is odd. The integers $q_1, q_2, \ldots, q_{n+1}$ are all odd positive integers less than $2n$. Because there are only n odd positive integers less than $2n$, it follows from the pigeonhole principle that two of the integers $q_1, q_2, \ldots, q_{n+1}$ must be equal. Therefore, there are distinct integers i and j such that $q_i = q_j$. Let q be the common value of q_i and q_j. Then, $a_i = 2^{k_i} q$ and $a_j = 2^{k_j} q$. It follows that if $k_i < k_j$, then a_i divides a_j; while if $k_i > k_j$, then a_j divides a_i. ◀

A clever application of the pigeonhole principle shows the existence of an increasing or a decreasing subsequence of a certain length in a sequence of distinct integers. We review some definitions before this application is presented. Suppose that $a_1, a_2, \ldots, a_N$ is a sequence of real numbers. A **subsequence** of this sequence is a sequence of the form $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$, where $1 \leq i_1 < i_2 < \cdots < i_m \leq N$. Hence, a subsequence is a sequence obtained from the original sequence by including some of the terms of the original sequence in their original order, and perhaps not including other terms. A sequence is called **strictly increasing** if each term is larger than the one that precedes it, and it is called **strictly decreasing** if each term is smaller than the one that precedes it.

THEOREM 3 Every sequence of $n^2 + 1$ distinct real numbers contains a subsequence of length $n + 1$ that is either strictly increasing or strictly decreasing.

We give an example before presenting the proof of Theorem 3.

EXAMPLE 12 The sequence 8, 11, 9, 1, 4, 6, 12, 10, 5, 7 contains 10 terms. Note that $10 = 3^2 + 1$. There are four strictly increasing subsequences of length four, namely, 1, 4, 6, 12; 1, 4, 6, 7; 1, 4, 6, 10; and 1, 4, 5, 7. There is also a strictly decreasing subsequence of length four, namely, 11, 9, 6, 5. ◀

The proof of the theorem will now be given.

Proof: Let $a_1, a_2, \ldots, a_{n^2+1}$ be a sequence of $n^2 + 1$ distinct real numbers. Associate an ordered pair with each term of the sequence, namely, associate (i_k, d_k) to the term a_k, where i_k is the length of the longest increasing subsequence starting at a_k, and d_k is the length of the longest decreasing subsequence starting at a_k.

Suppose that there are no increasing or decreasing subsequences of length $n + 1$. Then i_k and d_k are both positive integers less than or equal to n, for $k = 1, 2, \ldots, n^2 + 1$. Hence, by the product rule there are n^2 possible ordered pairs for (i_k, d_k). By the pigeonhole principle, two of these $n^2 + 1$ ordered pairs are equal. In other words, there exist terms a_s and a_t, with $s < t$ such that $i_s = i_t$ and $d_s = d_t$. We will show that this is impossible. Because the terms of the sequence are distinct, either $a_s < a_t$ or $a_s > a_t$. If $a_s < a_t$, then, because $i_s = i_t$, an increasing subsequence of length $i_t + 1$ can be built starting at a_s, by taking a_s followed by an increasing subsequence of length i_t beginning at a_t. This is a contradiction. Similarly, if $a_s > a_t$, the same reasoning shows that d_s must be greater than d_t, which is a contradiction. ◁

Links

The final example shows how the generalized pigeonhole principle can be applied to an important part of combinatorics called **Ramsey theory**, after the English mathematician F. P. Ramsey. In general, Ramsey theory deals with the distribution of subsets of elements of sets.

EXAMPLE 13 Assume that in a group of six people, each pair of individuals consists of two friends or two enemies. Show that there are either three mutual friends or three mutual enemies in the group.

Solution: Let A be one of the six people. Of the five other people in the group, there are either three or more who are friends of A, or three or more who are enemies of A. This follows from

Links

Courtesy of Stephen Frank Burch

FRANK PLUMPTON RAMSEY (1903–1930) Frank Plumpton Ramsey, son of the president of Magdalene College, Cambridge, was educated at Winchester and Trinity Colleges. After graduating in 1923, he was elected a fellow of King's College, Cambridge, where he spent the remainder of his life. Ramsey made important contributions to mathematical logic. What we now call Ramsey theory began with his clever combinatorial arguments, published in the paper "On a Problem of Formal Logic." Ramsey also made contributions to the mathematical theory of economics. He was noted as an excellent lecturer on the foundations of mathematics. According to one of his brothers, he was interested in almost everything, including English literature and politics. Ramsey was married and had two daughters. His death at the age of 26 resulting from chronic liver problems deprived the mathematical community and Cambridge University of a brilliant young scholar.

the generalized pigeonhole principle, because when five objects are divided into two sets, one of the sets has at least $\lceil 5/2 \rceil = 3$ elements. In the former case, suppose that B, C, and D are friends of A. If any two of these three individuals are friends, then these two and A form a group of three mutual friends. Otherwise, B, C, and D form a set of three mutual enemies. The proof in the latter case, when there are three or more enemies of A, proceeds in a similar manner. ◀

The **Ramsey number** $R(m, n)$, where m and n are positive integers greater than or equal to 2, denotes the minimum number of people at a party such that there are either m mutual friends or n mutual enemies, assuming that every pair of people at the party are friends or enemies. Example 13 shows that $R(3, 3) \leq 6$. We conclude that $R(3, 3) = 6$ because in a group of five people where every two people are friends or enemies, there may not be three mutual friends or three mutual enemies (see Exercise 28).

It is possible to prove some useful properties about Ramsey numbers, but for the most part it is difficult to find their exact values. Note that by symmetry it can be shown that $R(m, n) = R(n, m)$ (see Exercise 32). We also have $R(2, n) = n$ for every positive integer $n \geq 2$ (see Exercise 31). The exact values of only nine Ramsey numbers $R(m, n)$ with $3 \leq m \leq n$ are known, including $R(4, 4) = 18$. Only bounds are known for many other Ramsey numbers, including $R(5, 5)$, which is known to satisfy $43 \leq R(5, 5) \leq 49$. The reader interested in learning more about Ramsey numbers should consult [MiRo91] or [GrRoSp90].

Exercises

1. Show that in any set of six classes, each meeting regularly once a week on a particular day of the week, there must be two that meet on the same day, assuming that no classes are held on weekends.

2. Show that if there are 30 students in a class, then at least two have last names that begin with the same letter.

3. A drawer contains a dozen brown socks and a dozen black socks, all unmatched. A man takes socks out at random in the dark.

a) How many socks must he take out to be sure that he has at least two socks of the same color?

b) How many socks must he take out to be sure that he has at least two black socks?

4. A bowl contains 10 red balls and 10 blue balls. A woman selects balls at random without looking at them.

a) How many balls must she select to be sure of having at least three balls of the same color?

b) How many balls must she select to be sure of having at least three blue balls?

5. Undergraduate students at a college belong to one of four groups depending on the year in which they are expected to graduate. Each student must choose one of 21 different majors. How many students are needed to assure that there are two students expected to graduate in the same year who have the same major?

6. There are six professors teaching the introductory discrete mathematics class at a university. The same final exam is given by all six professors. If the lowest possible score on the final is 0 and the highest possible score is 100, how many students must there be to guarantee that there are two students with the same professor who earned the same final examination score?

7. Show that among any group of five (not necessarily consecutive) integers, there are two with the same remainder when divided by 4.

8. Let d be a positive integer. Show that among any group of $d + 1$ (not necessarily consecutive) integers there are two with exactly the same remainder when they are divided by d.

9. Let n be a positive integer. Show that in any set of n consecutive integers there is exactly one divisible by n.

10. Show that if f is a function from S to T, where S and T are finite sets with $|S| > |T|$, then there are elements s_1 and s_2 in S such that $f(s_1) = f(s_2)$, or in other words, f is not one-to-one.

11. What is the minimum number of students, each of whom comes from one of the 50 states, who must be enrolled in a university to guarantee that there are at least 100 who come from the same state?

***12.** Let (x_i, y_i), $i = 1, 2, 3, 4, 5$, be a set of five distinct points with integer coordinates in the xy plane. Show that the midpoint of the line joining at least one pair of these points has integer coordinates.

***13.** Let (x_i, y_i, z_i), $i = 1, 2, 3, 4, 5, 6, 7, 8, 9$, be a set of nine distinct points with integer coordinates in xyz space. Show that the midpoint of at least one pair of these points has integer coordinates.

14. How many ordered pairs of integers (a, b) are needed to guarantee that there are two ordered pairs (a_1, b_1) and (a_2, b_2) such that $a_1 \textbf{ mod } 5 = a_2 \textbf{ mod } 5$ and $b_1 \textbf{ mod } 5 = b_2 \textbf{ mod } 5$?

15. **a)** Show that if five integers are selected from the first eight positive integers, there must be a pair of these integers with a sum equal to 9.

b) Is the conclusion in part (a) true if four integers are selected rather than five?

16. **a)** Show that if seven integers are selected from the first 10 positive integers, there must be at least two pairs of these integers with the sum 11.

b) Is the conclusion in part (a) true if six integers are selected rather than seven?

17. How many numbers must be selected from the set $\{1, 2, 3, 4, 5, 6\}$ to guarantee that at least one pair of these numbers add up to 7?

18. How many numbers must be selected from the set $\{1, 3, 5, 7, 9, 11, 13, 15\}$ to guarantee that at least one pair of these numbers add up to 16?

19. A company stores products in a warehouse. Storage bins in this warehouse are specified by their aisle, location in the aisle, and shelf. There are 50 aisles, 85 horizontal locations in each aisle, and 5 shelves throughout the warehouse. What is the least number of products the company can have so that at least two products must be stored in the same bin?

20. Suppose that there are nine students in a discrete mathematics class at a small college.

a) Show that the class must have at least five male students or at least five female students.

b) Show that the class must have at least three male students or at least seven female students.

21. Suppose that every student in a discrete mathematics class of 25 students is a freshman, a sophomore, or a junior.

a) Show that there are at least nine freshmen, at least nine sophomores, or at least nine juniors in the class.

b) Show that there are either at least three freshmen, at least 19 sophomores, or at least five juniors in the class.

22. Find an increasing subsequence of maximal length and a decreasing subsequence of maximal length in the sequence 22, 5, 7, 2, 23, 10, 15, 21, 3, 17.

23. Construct a sequence of 16 positive integers that has no increasing or decreasing subsequence of five terms.

24. Show that if there are 101 people of different heights standing in a line, it is possible to find 11 people in the order they are standing in the line with heights that are either increasing or decreasing.

***25.** Show that whenever 25 girls and 25 boys are seated around a circular table there is always a person both of whose neighbors are boys.

****26.** Suppose that 21 girls and 21 boys enter a mathematics competition. Furthermore, suppose that each entrant solves at most six questions, and for every boy-girl pair, there is at least one question that they both solved. Show that there is a question that was solved by at least three girls and at least three boys.

***27.** Describe an algorithm in pseudocode for producing the largest increasing or decreasing subsequence of a sequence of distinct integers.

28. Show that in a group of five people (where any two people are either friends or enemies), there are not necessarily three mutual friends or three mutual enemies.

29. Show that in a group of 10 people (where any two people are either friends or enemies), there are either three mutual friends or four mutual enemies, and there are either three mutual enemies or four mutual friends.

30. Use Exercise 29 to show that among any group of 20 people (where any two people are either friends or enemies), there are either four mutual friends or four mutual enemies.

31. Show that if n is an integer with $n \geq 2$, then the Ramsey number $R(2, n)$ equals n. (Recall that Ramsey numbers were discussed after Example 13 in Section 6.2.)

32. Show that if m and n are integers with $m \geq 2$ and $n \geq 2$, then the Ramsey numbers $R(m, n)$ and $R(n, m)$ are equal. (Recall that Ramsey numbers were discussed after Example 13 in Section 6.2.)

33. Show that there are at least six people in California (population: 39 million) with the same three initials who were born on the same day of the year (but not necessarily in the same year). Assume that everyone has three initials.

34. Show that if there are 100,000,000 wage earners in the United States who earn less than 1,000,000 dollars (but at least a penny), then there are two who earned exactly the same amount of money, to the penny, last year.

35. In the 17th century, there were more than 800,000 inhabitants of Paris. At the time, it was believed that no one had more than 200,000 hairs on their head. Assuming these numbers are correct and that everyone has at least one hair on their head (that is, no one is completely bald), use the pigeonhole principle to show, as the French writer Pierre Nicole did, that there had to be two Parisians with the same number of hairs on their heads. Then use the generalized pigeonhole principle to show that there had to be at least five Parisians at that time with the same number of hairs on their heads.

36. Assuming that no one has more than 1,000,000 hairs on their head and that the population of New York City was 8,537,673 in 2016, show there had to be at least nine people in New York City in 2016 with the same number of hairs on their heads.

37. There are 38 different time periods during which classes at a university can be scheduled. If there are 677 different classes, how many different rooms will be needed?

38. A computer network consists of six computers. Each computer is directly connected to at least one of the other computers. Show that there are at least two computers in the network that are directly connected to the same number of other computers.

39. A computer network consists of six computers. Each computer is directly connected to zero or more of the other computers. Show that there are at least two computers in the network that are directly connected to the same number of other computers. [*Hint:* It is impossible to have a computer linked to none of the others and a computer linked to all the others.]

40. Find the least number of cables required to connect eight computers to four printers to guarantee that for every choice of four of the eight computers, these four computers can directly access four different printers. Justify your answer.

41. Find the least number of cables required to connect 100 computers to 20 printers to guarantee that every subset of 20 computers can directly access 20 different printers. (Here, the assumptions about cables and computers are the same as in Example 9.) Justify your answer.

***42.** Prove that at a party where there are at least two people, there are two people who know the same number of other people there.

43. An arm wrestler is the champion for a period of 75 hours. (Here, by an hour, we mean a period starting from an exact hour, such as 1 P.M., until the next hour.) The arm wrestler had at least one match an hour, but no more than 125 total matches. Show that there is a period of consecutive hours during which the arm wrestler had exactly 24 matches.

***44.** Is the statement in Exercise 43 true if 24 is replaced by

a) 2? **b)** 23? **c)** 25? **d)** 30?

45. Show that if f is a function from S to T, where S and T are nonempty finite sets and $m = \lceil |S| / |T| \rceil$, then there are at least m elements of S mapped to the same value of T. That is, show that there are distinct elements $s_1, s_2, \ldots, s_m$ of S such that $f(s_1) = f(s_2) = \cdots = f(s_m)$.

46. There are 51 houses on a street. Each house has an address between 1000 and 1099, inclusive. Show that at least two houses have addresses that are consecutive integers.

***47.** Let x be an irrational number. Show that for some positive integer j not exceeding the positive integer n, the absolute value of the difference between jx and the nearest integer to jx is less than $1/n$.

48. Let $n_1, n_2, \ldots, n_t$ be positive integers. Show that if $n_1 + n_2 + \cdots + n_t - t + 1$ objects are placed into t boxes, then for some i, $i = 1, 2, \ldots, t$, the ith box contains at least n_i objects.

***49.** An alternative proof of Theorem 3 based on the generalized pigeonhole principle is outlined in this exercise. The notation used is the same as that used in the proof in the text.

a) Assume that $i_k \leq n$ for $k = 1, 2, \ldots, n^2 + 1$. Use the generalized pigeonhole principle to show that there are $n + 1$ terms $a_{k_1}, a_{k_2}, \ldots, a_{k_{n+1}}$ with $i_{k_1} = i_{k_2} = \cdots = i_{k_{n+1}}$, where $1 \leq k_1 < k_2 < \cdots < k_{n+1}$.

b) Show that $a_{k_j} > a_{k_{j+1}}$ for $j = 1, 2, \ldots, n$. [*Hint:* Assume that $a_{k_j} < a_{k_{j+1}}$, and show that this implies that $i_{k_j} > i_{k_{j+1}}$, which is a contradiction.]

c) Use parts (a) and (b) to show that if there is no increasing subsequence of length $n + 1$, then there must be a decreasing subsequence of this length.

6.3 Permutations and Combinations

6.3.1 Introduction

Many counting problems can be solved by finding the number of ways to arrange a specified number of distinct elements of a set of a particular size, where the order of these elements matters. Many other counting problems can be solved by finding the number of ways to select a particular number of elements from a set of a particular size, where the order of the elements selected does not matter. For example, in how many ways can we select three students from a group of five students to stand in line for a picture? How many different committees of three students can be formed from a group of four students? In this section we will develop methods to answer questions such as these.

6.3.2 Permutations

We begin by solving the first question posed in the introduction to this section, as well as related questions.

EXAMPLE 1 In how many ways can we select three students from a group of five students to stand in line for a picture? In how many ways can we arrange all five of these students in a line for a picture?

Extra Examples

Solution: First, note that the order in which we select the students matters. There are five ways to select the first student to stand at the start of the line. Once this student has been selected, there are four ways to select the second student in the line. After the first and second students have been selected, there are three ways to select the third student in the line. By the product rule, there are $5 \cdot 4 \cdot 3 = 60$ ways to select three students from a group of five students to stand in line for a picture.

To arrange all five students in a line for a picture, we select the first student in five ways, the second in four ways, the third in three ways, the fourth in two ways, and the fifth in one way. Consequently, there are $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ ways to arrange all five students in a line for a picture. ◀

Example 1 illustrates how ordered arrangements of distinct objects can be counted. This leads to some terminology.

Links

A **permutation** of a set of distinct objects is an ordered arrangement of these objects. We also are interested in ordered arrangements of some of the elements of a set. An ordered arrangement of r elements of a set is called an ***r*-permutation**.

EXAMPLE 2 Let $S = \{1, 2, 3\}$. The ordered arrangement 3, 1, 2 is a permutation of S. The ordered arrangement 3, 2 is a 2-permutation of S. ◀

The number of r-permutations of a set with n elements is denoted by $P(n, r)$. We can find $P(n, r)$ using the product rule.

EXAMPLE 3 Let $S = \{a, b, c\}$. The 2-permutations of S are the ordered arrangements a, b; a, c; b, a; b, c; c, a; and c, b. Consequently, there are six 2-permutations of this set with three elements. There are always six 2-permutations of a set with three elements. There are three ways to choose the first element of the arrangement. There are two ways to choose the second element of the arrangement, because it must be different from the first element. Hence, by the product rule, we see that $P(3, 2) = 3 \cdot 2 = 6$. the first element. By the product rule, it follows that $P(3, 2) = 3 \cdot 2 = 6$. ◀

We now use the product rule to find a formula for $P(n, r)$ whenever n and r are positive integers with $1 \le r \le n$.

THEOREM 1 If n is a positive integer and r is an integer with $1 \le r \le n$, then there are

$$P(n, r) = n(n-1)(n-2)\cdots(n-r+1)$$

r-permutations of a set with n distinct elements.

Proof: We will use the product rule to prove that this formula is correct. The first element of the permutation can be chosen in n ways because there are n elements in the set. There are $n-1$ ways to choose the second element of the permutation, because there are $n-1$ elements left in the set after using the element picked for the first position. Similarly, there are $n-2$ ways to choose the third element, and so on, until there are exactly $n-(r-1) = n-r+1$ ways to choose the rth element. Consequently, by the product rule, there are

$$n(n-1)(n-2)\cdots(n-r+1)$$

r-permutations of the set. ◁

Note that $P(n, 0) = 1$ whenever n is a nonnegative integer because there is exactly one way to order zero elements. That is, there is exactly one list with no elements in it, namely the empty list.

We now state a useful corollary of Theorem 1.

COROLLARY 1 If n and r are integers with $0 \leq r \leq n$, then $P(n, r) = \dfrac{n!}{(n-r)!}$.

Proof: When n and r are integers with $1 \leq r \leq n$, by Theorem 1 we have

$$P(n, r) = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

Because $\dfrac{n!}{(n-0)!} = \dfrac{n!}{n!} = 1$ whenever n is a nonnegative integer, we see that the formula $P(n, r) = \dfrac{n!}{(n-r)!}$ also holds when $r = 0$. ◁

By Theorem 1 we know that if n is a positive integer, then $P(n, n) = n!$. We will illustrate this result with some examples.

EXAMPLE 4 How many ways are there to select a first-prize winner, a second-prize winner, and a third-prize winner from 100 different people who have entered a contest?

Solution: Because it matters which person wins which prize, the number of ways to pick the three prize winners is the number of ordered selections of three elements from a set of 100 elements, that is, the number of 3-permutations of a set of 100 elements. Consequently, the answer is

$$P(100, 3) = 100 \cdot 99 \cdot 98 = 970{,}200.$$ ◀

EXAMPLE 5 Suppose that there are eight runners in a race. The winner receives a gold medal, the second-place finisher receives a silver medal, and the third-place finisher receives a bronze medal. How many different ways are there to award these medals, if all possible outcomes of the race can occur and there are no ties?

Solution: The number of different ways to award the medals is the number of 3-permutations of a set with eight elements. Hence, there are $P(8, 3) = 8 \cdot 7 \cdot 6 = 336$ possible ways to award the medals. ◀

EXAMPLE 6 Suppose that a saleswoman has to visit eight different cities. She must begin her trip in a specified city, but she can visit the other seven cities in any order she wishes. How many possible orders can the saleswoman use when visiting these cities?

Solution: The number of possible paths between the cities is the number of permutations of seven elements, because the first city is determined, but the remaining seven can be ordered arbitrarily. Consequently, there are $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$ ways for the saleswoman to choose her tour. If, for instance, the saleswoman wishes to find the path between the cities with minimum distance, and she computes the total distance for each possible path, she must consider a total of 5040 paths! ◀

EXAMPLE 7 How many permutations of the letters $ABCDEFGH$ contain the string ABC?

Solution: Because the letters ABC must occur as a block, we can find the answer by finding the number of permutations of six objects, namely, the block ABC and the individual letters D, E, F, G, and H. Because these six objects can occur in any order, there are $6! = 720$ permutations of the letters $ABCDEFGH$ in which ABC occurs as a block. ◀

6.3.3 Combinations

We now turn our attention to counting unordered selections of objects. We begin by solving a question posed in the introduction to this section of the chapter.

EXAMPLE 8 How many different committees of three students can be formed from a group of four students?

Solution: To answer this question, we need only find the number of subsets with three elements from the set containing the four students. We see that there are four such subsets, one for each of the four students, because choosing three students is the same as choosing one of the four students to leave out of the group. This means that there are four ways to choose the three students for the committee, where the order in which these students are chosen does not matter. ◀

Links

Example 8 illustrates that many counting problems can be solved by finding the number of subsets of a particular size of a set with n elements, where n is a positive integer.

An ***r*-combination** of elements of a set is an unordered selection of r elements from the set. Thus, an r-combination is simply a subset of the set with r elements.

EXAMPLE 9 Let S be the set $\{1, 2, 3, 4\}$. Then $\{1, 3, 4\}$ is a 3-combination from S. (Note that $\{4, 1, 3\}$ is the same 3-combination as $\{1, 3, 4\}$, because the order in which the elements of a set are listed does not matter.) ◀

The number of r-combinations of a set with n distinct elements is denoted by $C(n, r)$. Note that $C(n, r)$ is also denoted by $\binom{n}{r}$ and is called a **binomial coefficient**. We will learn where this terminology comes from in Section 6.4.

EXAMPLE 10 We see that $C(4, 2) = 6$, because the 2-combinations of $\{a, b, c, d\}$ are the six subsets $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$, and $\{c, d\}$. ◀

We can determine the number of r-combinations of a set with n elements using the formula for the number of r-permutations of a set. To do this, note that the r-permutations of a set can be obtained by first forming r-combinations and then ordering the elements in these combinations. The proof of Theorem 2, which gives the value of $C(n, r)$, is based on this observation.

THEOREM 2 The number of r-combinations of a set with n elements, where n is a nonnegative integer and r is an integer with $0 \le r \le n$, equals

$$C(n, r) = \frac{n!}{r!\,(n-r)!}.$$

Proof: The $P(n, r)$ r-permutations of the set can be obtained by forming the $C(n, r)$ r-combinations of the set, and then ordering the elements in each r-combination, which can be done in $P(r, r)$ ways. Consequently, by the product rule,

$$P(n, r) = C(n, r) \cdot P(r, r).$$

This implies that

$$C(n, r) = \frac{P(n, r)}{P(r, r)} = \frac{n!/(n-r)!}{r!/(r-r)!} = \frac{n!}{r!\,(n-r)!}.$$

We can also use the division rule for counting to construct a proof of this theorem. Because the order of elements in a combination does not matter and there are $P(r, r)$ ways to order r elements in an r-combination of n elements, each of the $C(n, r)$ r-combinations of a set with n elements corresponds to exactly $P(r, r)$ r-permutations. Hence, by the division rule, $C(n, r) = \frac{P(n,r)}{P(r,r)}$, which implies as before that $C(n, r) = \frac{n!}{r!\,(n-r)!}$. ◁

The formula in Theorem 2, although explicit, is not helpful when $C(n, r)$ is computed for large values of n and r. The reasons are that it is practical to compute exact values of factorials exactly only for small integer values, and when floating point arithmetic is used, the formula in Theorem 2 may produce a value that is not an integer. When computing $C(n, r)$, first note that when we cancel out $(n - r)!$ from the numerator and denominator of the expression for $C(n, r)$ in Theorem 2, we obtain

$$C(n, r) = \frac{n!}{r!\,(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!}.$$

Consequently, to compute $C(n, r)$ you can cancel out all the terms in the larger factorial in the denominator from the numerator and denominator, then multiply all the terms that do not cancel in the numerator and finally divide by the smaller factorial in the denominator. [When doing this calculation by hand, instead of by machine, it is also worthwhile to factor out common factors in the numerator $n(n-1)\cdots(n-r+1)$ and in the denominator $r!$.] Note that many computational programs can be used to find $C(n, r)$. [Such functions may be called *choose*(n, k) or *binom*(n, k).]

Example 11 illustrates how $C(n, k)$ is computed when k is relatively small compared to n and when k is close to n. It also illustrates a key identity enjoyed by the numbers $C(n, k)$.

EXAMPLE 11 How many poker hands of five cards can be dealt from a standard deck of 52 cards? Also, how many ways are there to select 47 cards from a standard deck of 52 cards?

Solution: Because the order in which the five cards are dealt from a deck of 52 cards does not matter, there are

$$C(52, 5) = \frac{52!}{5!47!}$$

different hands of five cards that can be dealt. To compute the value of $C(52, 5)$, first divide the numerator and denominator by 47! to obtain

$$C(52, 5) = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}.$$

This expression can be simplified by first dividing the factor 5 in the denominator into the factor 50 in the numerator to obtain a factor 10 in the numerator, then dividing the factor 4 in the denominator into the factor 48 in the numerator to obtain a factor of 12 in the numerator, then dividing the factor 3 in the denominator into the factor 51 in the numerator to obtain a factor of 17 in the numerator, and finally, dividing the factor 2 in the denominator into the factor 52 in the numerator to obtain a factor of 26 in the numerator. We find that

$$C(52, 5) = 26 \cdot 17 \cdot 10 \cdot 49 \cdot 12 = 2{,}598{,}960.$$

Consequently, there are 2,598,960 different poker hands of five cards that can be dealt from a standard deck of 52 cards.

Note that there are

$$C(52, 47) = \frac{52!}{47!5!}$$

different ways to select 47 cards from a standard deck of 52 cards. We do not need to compute this value because $C(52, 47) = C(52, 5)$. (Only the order of the factors 5! and 47! is different in the denominators in the formulae for these quantities.) It follows that there are also 2,598,960 different ways to select 47 cards from a standard deck of 52 cards. ◀

In Example 11 we observed that $C(52, 5) = C(52, 47)$. This is not surprising because selecting five cards out of 52 is the same as selecting the 47 that we leave out. The identity $C(52, 5) = C(52, 47)$ is a special case of the useful identity for the number of r-combinations of a set given in Corollary 2.

COROLLARY 2 Let n and r be nonnegative integers with $r \leq n$. Then $C(n, r) = C(n, n - r)$.

Proof: From Theorem 2 it follows that

$$C(n, r) = \frac{n!}{r!\,(n-r)!}$$

and

$$C(n, n-r) = \frac{n!}{(n-r)!\,[n-(n-r)]!} = \frac{n!}{(n-r)!\,r!}.$$

Hence, $C(n, r) = C(n, n - r)$. ◁

We can also prove Corollary 2 without relying on algebraic manipulation. Instead, we can use a combinatorial proof. We describe this important type of proof in Definition 1.

Definition 1 A *combinatorial proof* of an identity is a proof that uses counting arguments to prove that both sides of the identity count the same objects but in different ways or a proof that is based on showing that there is a bijection between the sets of objects counted by the two sides of the identity. These two types of proofs are called *double counting proofs* and *bijective proofs*, respectively.

Combinatorial proofs are almost always much shorter and provide more insights than proofs based on algebraic manipulation.

Many identities involving binomial coefficients can be proved using combinatorial proofs. We now show how to prove Corollary 2 using a combinatorial proof. We will provide both a double counting proof and a bijective proof, both based on the same basic idea.

Proof: We will use a bijective proof to show that $C(n, r) = C(n, n - r)$ for all integers n and r with $0 \leq r \leq n$. Suppose that S is a set with n elements. The function that maps a subset A of S to $\overline{A}$ is a bijection between subsets of S with r elements and subsets with $n - r$ elements (as the reader should verify). The identity $C(n, r) = C(n, n - r)$ follows because when there is a bijection between two finite sets, the two sets must have the same number of elements.

Alternatively, we can reformulate this argument as a double counting proof. By definition, the number of subsets of S with r elements equals $C(n, r)$. But each subset A of S is also determined by specifying which elements are not in A, and so are in $\overline{A}$. Because the complement of a subset of S with r elements has $n - r$ elements, there are also $C(n, n - r)$ subsets of S with r elements. It follows that $C(n, r) = C(n, n - r)$. ◁

EXAMPLE 12 How many ways are there to select five players from a 10-member tennis team to make a trip to a match at another school?

Extra Examples

Solution: The answer is given by the number of 5-combinations of a set with 10 elements. By Theorem 2, the number of such combinations is

$$C(10, 5) = \frac{10!}{5!\,5!} = 252.$$

◀

EXAMPLE 13 A group of 30 people have been trained as astronauts to go on the first mission to Mars. How many ways are there to select a crew of six people to go on this mission (assuming that all crew members have the same job)?

Solution: The number of ways to select a crew of six from the pool of 30 people is the number of 6-combinations of a set with 30 elements, because the order in which these people are chosen does not matter. By Theorem 2, the number of such combinations is

$$C(30, 6) = \frac{30!}{6!\,24!} = \frac{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 593{,}775.$$

◀

EXAMPLE 14 How many bit strings of length n contain exactly r 1s?

Solution: The positions of r 1s in a bit string of length n form an r-combination of the set $\{1, 2, 3, \ldots, n\}$. Hence, there are $C(n, r)$ bit strings of length n that contain exactly r 1s. ◀

EXAMPLE 15 Suppose that there are 9 faculty members in the mathematics department and 11 in the computer science department. How many ways are there to select a committee to develop a discrete mathematics course at a school if the committee is to consist of three faculty members from the mathematics department and four from the computer science department?

Solution: By the product rule, the answer is the product of the number of 3-combinations of a set with nine elements and the number of 4-combinations of a set with 11 elements. By Theorem 2, the number of ways to select the committee is

$$C(9, 3) \cdot C(11, 4) = \frac{9!}{3!6!} \cdot \frac{11!}{4!7!} = 84 \cdot 330 = 27{,}720.$$

◀

Exercises

1. List all the permutations of $\{a, b, c\}$.
2. How many different permutations are there of the set $\{a, b, c, d, e, f, g\}$?
3. How many permutations of $\{a, b, c, d, e, f, g\}$ end with a?
4. Let $S = \{1, 2, 3, 4, 5\}$.
   a) List all the 3-permutations of S.
   b) List all the 3-combinations of S.
5. Find the value of each of these quantities.
   a) $P(6, 3)$ b) $P(6, 5)$
   c) $P(8, 1)$ d) $P(8, 5)$
   e) $P(8, 8)$ f) $P(10, 9)$
6. Find the value of each of these quantities.
   a) $C(5, 1)$ b) $C(5, 3)$
   c) $C(8, 4)$ d) $C(8, 8)$
   e) $C(8, 0)$ f) $C(12, 6)$
7. Find the number of 5-permutations of a set with nine elements.
8. In how many different orders can five runners finish a race if no ties are allowed?
9. How many possibilities are there for the win, place, and show (first, second, and third) positions in a horse race with 12 horses if all orders of finish are possible?
10. There are six different candidates for governor of a state. In how many different orders can the names of the candidates be printed on a ballot?
11. How many bit strings of length 10 contain
    a) exactly four 1s?
    b) at most four 1s?
    c) at least four 1s?
    d) an equal number of 0s and 1s?
12. How many bit strings of length 12 contain
    a) exactly three 1s?
    b) at most three 1s?
    c) at least three 1s?
    d) an equal number of 0s and 1s?
13. A group contains n men and n women. How many ways are there to arrange these people in a row if the men and women alternate?
14. In how many ways can a set of two positive integers less than 100 be chosen?
15. In how many ways can a set of five letters be selected from the English alphabet?
16. How many subsets with an odd number of elements does a set with 10 elements have?
17. How many subsets with more than two elements does a set with 100 elements have?
18. A coin is flipped eight times where each flip comes up either heads or tails. How many possible outcomes
    a) are there in total?
    b) contain exactly three heads?
    c) contain at least three heads?
    d) contain the same number of heads and tails?
19. A coin is flipped 10 times where each flip comes up either heads or tails. How many possible outcomes
    a) are there in total?
    b) contain exactly two heads?
    c) contain at most three tails?
    d) contain the same number of heads and tails?
20. How many bit strings of length 10 have
    a) exactly three 0s?
    b) more 0s than 1s?
    c) at least seven 1s?
    d) at least three 1s?
21. How many permutations of the letters *ABCDEFG* contain
    a) the string *BCD*?
    b) the string *CFGA*?
    c) the strings *BA* and *GF*?
    d) the strings *ABC* and *DE*?
    e) the strings *ABC* and *CDE*?
    f) the strings *CBA* and *BED*?
22. How many permutations of the letters *ABCDEFGH* contain
    a) the string *ED*?
    b) the string *CDE*?
    c) the strings *BA* and *FGH*?
    d) the strings *AB*, *DE*, and *GH*?
    e) the strings *CAB* and *BED*?
    f) the strings *BCA* and *ABF*?
23. How many ways are there for eight men and five women to stand in a line so that no two women stand next to each other? [*Hint:* First position the men and then consider possible positions for the women.]
24. How many ways are there for 10 women and six men to stand in a line so that no two men stand next to each other? [*Hint:* First position the women and then consider possible positions for the men.]
25. How many ways are there for four men and five women to stand in a line so that
    a) all men stand together?
    b) all women stand together?
26. How many ways are there for three penguins and six puffins to stand in a line so that
    a) all puffins stand together?
    b) all penguins stand together?
27. One hundred tickets, numbered $1, 2, 3, \ldots, 100$, are sold to 100 different people for a drawing. Four different prizes are awarded, including a grand prize (a trip to Tahiti). How many ways are there to award the prizes if
    a) there are no restrictions?
    b) the person holding ticket 47 wins the grand prize?
    c) the person holding ticket 47 wins one of the prizes?
    d) the person holding ticket 47 does not win a prize?
    e) the people holding tickets 19 and 47 both win prizes?
    f) the people holding tickets 19, 47, and 73 all win prizes?

g) the people holding tickets 19, 47, 73, and 97 all win prizes?
h) none of the people holding tickets 19, 47, 73, and 97 wins a prize?
i) the grand prize winner is a person holding ticket 19, 47, 73, or 97?
j) the people holding tickets 19 and 47 win prizes, but the people holding tickets 73 and 97 do not win prizes?

28. Thirteen people on a softball team show up for a game.
a) How many ways are there to choose 10 players to take the field?
b) How many ways are there to assign the 10 positions by selecting players from the 13 people who show up?
c) Of the 13 people who show up, three are women. How many ways are there to choose 10 players to take the field if at least one of these players must be a woman?

29. A club has 25 members.
a) How many ways are there to choose four members of the club to serve on an executive committee?
b) How many ways are there to choose a president, vice president, secretary, and treasurer of the club, where no person can hold more than one office?

30. A professor writes 40 discrete mathematics true/false questions. Of the statements in these questions, 17 are true. If the questions can be positioned in any order, how many different answer keys are possible?

***31.** How many 4-permutations of the positive integers not exceeding 100 contain three consecutive integers k, $k+1$, $k+2$, in the correct order
a) where these consecutive integers can perhaps be separated by other integers in the permutation?
b) where they are in consecutive positions in the permutation?

32. Seven women and nine men are on the faculty in the mathematics department at a school.
a) How many ways are there to select a committee of five members of the department if at least one woman must be on the committee?
b) How many ways are there to select a committee of five members of the department if at least one woman and at least one man must be on the committee?

33. The English alphabet contains 21 consonants and five vowels. How many strings of six lowercase letters of the English alphabet contain
a) exactly one vowel?
b) exactly two vowels?
c) at least one vowel?
d) at least two vowels?

34. How many strings of six lowercase letters from the English alphabet contain
a) the letter a?
b) the letters a and b?
c) the letters a and b in consecutive positions with a preceding b, with all the letters distinct?
d) the letters a and b, where a is somewhere to the left of b in the string, with all the letters distinct?

35. Suppose that a department contains 10 men and 15 women. How many ways are there to form a committee with six members if it must have the same number of men and women?

36. Suppose that a department contains 10 men and 15 women. How many ways are there to form a committee with six members if it must have more women than men?

37. How many bit strings contain exactly eight 0s and 10 1s if every 0 must be immediately followed by a 1?

38. How many bit strings contain exactly five 0s and 14 1s if every 0 must be immediately followed by two 1s?

39. How many bit strings of length 10 contain at least three 1s and at least three 0s?

40. How many ways are there to select 12 countries in the United Nations to serve on a council if 3 are selected from a block of 45, 4 are selected from a block of 57, and the others are selected from the remaining 69 countries?

41. How many license plates consisting of three letters followed by three digits contain no letter or digit twice?

A **circular *r*-permutation of *n*** people is a seating of r of these n people around a circular table, where seatings are considered to be the same if they can be obtained from each other by rotating the table.

42. Find the number of circular 3-permutations of 5 people.

43. Find a formula for the number of circular r-permutations of n people.

44. Find a formula for the number of ways to seat r of n people around a circular table, where seatings are considered the same if every person has the same two neighbors without regard to which side these neighbors are sitting on.

45. How many ways are there for a horse race with three horses to finish if ties are possible? [*Note:* Two or three horses may tie.]

***46.** How many ways are there for a horse race with four horses to finish if ties are possible? [*Note:* Any number of the four horses may tie.]

***47.** There are six runners in the 100-yard dash. How many ways are there for three medals to be awarded if ties are possible? (The runner or runners who finish with the fastest time receive gold medals, the runner or runners who finish with exactly one runner ahead receive silver medals, and the runner or runners who finish with exactly two runners ahead receive bronze medals.)

***48.** This procedure is used to break ties in games in the championship round of the World Cup soccer tournament. Each team selects five players in a prescribed order. Each of these players takes a penalty kick, with a player from the first team followed by a player from the second team and so on, following the order of players specified. If the score is still tied at the end of the 10 penalty kicks, this procedure is repeated. If the score is still tied after 20 penalty kicks, a sudden-death shootout occurs, with the first team scoring an unanswered goal victorious.

a) How many different scoring scenarios are possible if the game is settled in the first round of 10 penalty kicks, where the round ends once it is impossible for a team to equal the number of goals scored by the other team?

b) How many different scoring scenarios for the first and second groups of penalty kicks are possible if the game is settled in the second round of 10 penalty kicks?

c) How many scoring scenarios are possible for the full set of penalty kicks if the game is settled with no more than 10 total additional kicks after the two rounds of five kicks for each team?

6.4 Binomial Coefficients and Identities

As we remarked in Section 6.3, the number of r-combinations from a set with n elements is often denoted by $\binom{n}{r}$. This number is also called a **binomial coefficient** because these numbers occur as coefficients in the expansion of powers of binomial expressions such as $(a+b)^n$. We will discuss the **binomial theorem**, which gives a power of a binomial expression as a sum of terms involving binomial coefficients. We will prove this theorem using a combinatorial proof. We will also show how combinatorial proofs can be used to establish some of the many different identities that express relationships among binomial coefficients.

6.4.1 The Binomial Theorem

Links

The binomial theorem gives the coefficients of the expansion of powers of binomial expressions. A **binomial** expression is simply the sum of two terms, such as $x+y$. (The terms can be products of constants and variables, but that does not concern us here.)

Example 1 illustrates how the coefficients in a typical expansion can be found and prepares us for the statement of the binomial theorem.

EXAMPLE 1 The expansion of $(x+y)^3$ can be found using combinatorial reasoning instead of multiplying the three terms out. When $(x+y)^3 = (x+y)(x+y)(x+y)$ is expanded, all products of a term in the first sum, a term in the second sum, and a term in the third sum are added. Terms of the form x^3, x^2y, xy^2, and y^3 arise. To obtain a term of the form x^3, an x must be chosen in each of the sums, and this can be done in only one way. Thus, the x^3 term in the product has a coefficient of 1. To obtain a term of the form x^2y, an x must be chosen in two of the three sums (and consequently a y in the other sum). Hence, the number of such terms is the number of 2-combinations of three objects, namely, $\binom{3}{2}$. Similarly, the number of terms of the form xy^2 is the number of ways to pick one of the three sums to obtain an x (and consequently take a y from each of the other two sums). This can be done in $\binom{3}{1}$ ways. Finally, the only way to obtain a y^3 term is to choose the y for each of the three sums in the product, and this can be done in exactly one way. Consequently, it follows that

$$\begin{aligned}(x+y)^3 &= (x+y)(x+y)(x+y) = (xx+xy+yx+yy)(x+y)\\ &= xxx+xxy+xyx+xyy+yxx+yxy+yyx+yyy\\ &= x^3+3x^2y+3xy^2+y^3.\end{aligned}$$

◄

We now state the binomial theorem.

THEOREM 1 **THE BINOMIAL THEOREM** Let x and y be variables, and let n be a nonnegative integer. Then

$$(x+y)^n = \sum_{j=0}^{n}\binom{n}{j}x^{n-j}y^j = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n.$$

Proof: We use a combinatorial proof. The terms in the product when it is expanded are of the form $x^{n-j}y^j$ for $j = 0, 1, 2, \ldots, n$. To count the number of terms of the form $x^{n-j}y^j$, note that to obtain such a term it is necessary to choose $n - j$ xs from the n binomial factors (so that the other j terms in the product are ys). Therefore, the coefficient of $x^{n-j}y^j$ is $\binom{n}{n-j}$, which is equal to $\binom{n}{j}$. This proves the theorem. ◁

Some computational uses of the binomial theorem are illustrated in Examples 2–4.

EXAMPLE 2 What is the expansion of $(x + y)^4$?

Extra Examples

Solution: From the binomial theorem it follows that

$$\begin{aligned}(x+y)^4 &= \sum_{j=0}^{4}\binom{4}{j}x^{4-j}y^j\\ &= \binom{4}{0}x^4 + \binom{4}{1}x^3y + \binom{4}{2}x^2y^2 + \binom{4}{3}xy^3 + \binom{4}{4}y^4\\ &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.\end{aligned}$$

◀

EXAMPLE 3 What is the coefficient of $x^{12}y^{13}$ in the expansion of $(x + y)^{25}$?

Solution: From the binomial theorem it follows that this coefficient is

$$\binom{25}{13} = \frac{25!}{13!\,12!} = 5{,}200{,}300.$$

◀

EXAMPLE 4 What is the coefficient of $x^{12}y^{13}$ in the expansion of $(2x - 3y)^{25}$?

Solution: First, note that this expression equals $(2x + (-3y))^{25}$. By the binomial theorem, we have

$$(2x + (-3y))^{25} = \sum_{j=0}^{25}\binom{25}{j}(2x)^{25-j}(-3y)^j.$$

Consequently, the coefficient of $x^{12}y^{13}$ in the expansion is obtained when $j = 13$, namely,

$$\binom{25}{13}2^{12}(-3)^{13} = -\frac{25!}{13!\,12!}2^{12}3^{13}.$$

Note that another way to find the solution is to first use the binomial theorem to see that

$$(u + v)^{25} = \sum_{j=0}^{25}\binom{25}{j}u^{25-j}v^j.$$

Setting $u = 2x$ and $v = -3y$ in this equation yields the same result. ◀

We can prove some useful identities using the binomial theorem, as Corollaries 1, 2, and 3 demonstrate.

COROLLARY 1 Let n be a nonnegative integer. Then

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

Proof: Using the binomial theorem with $x = 1$ and $y = 1$, we see that

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

This is the desired result. ◁

There is also a nice combinatorial proof of Corollary 1, which we now present.

Proof: A set with n elements has a total of 2^n different subsets. Each subset has zero elements, one element, two elements, …, or n elements in it. There are $\binom{n}{0}$ subsets with zero elements, $\binom{n}{1}$ subsets with one element, $\binom{n}{2}$ subsets with two elements, …, and $\binom{n}{n}$ subsets with n elements. Therefore,

$$\sum_{k=0}^{n} \binom{n}{k}$$

counts the total number of subsets of a set with n elements. By equating the two formulas we have for the number of subsets of a set with n elements, we see that

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

◁

COROLLARY 2 Let n be a positive integer. Then

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

Proof: When we use the binomial theorem with $x = -1$ and $y = 1$, we see that

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

This proves the corollary. ◁

Remark: Corollary 2 implies that

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots.$$

COROLLARY 3 Let n be a nonnegative integer. Then

$$\sum_{k=0}^{n} 2^k \binom{n}{k} = 3^n.$$

Proof: We recognize that the left-hand side of this formula is the expansion of $(1+2)^n$ provided by the binomial theorem. Therefore, by the binomial theorem, we see that

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

Hence

$$\sum_{k=0}^{n} 2^k \binom{n}{k} = 3^n.$$

◁

6.4.2 Pascal's Identity and Triangle

The binomial coefficients satisfy many different identities. We introduce one of the most important of these now.

THEOREM 2 **PASCAL'S IDENTITY** Let n and k be positive integers with $n \geq k$. Then

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

Proof: We will use a combinatorial proof. Suppose that T is a set containing $n+1$ elements. Let a be an element in T, and let $S = T - \{a\}$. Note that there are $\binom{n+1}{k}$ subsets of T containing k elements. However, a subset of T with k elements either contains a together with $k-1$ elements of S, or contains k elements of S and does not contain a. Because there are $\binom{n}{k-1}$ subsets of $k-1$ elements of S, there are $\binom{n}{k-1}$ subsets of k elements of T that contain a. And there are $\binom{n}{k}$ subsets of k elements of T that do not contain a, because there are $\binom{n}{k}$ subsets of k elements of S. Consequently,

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

◁

Remark: It is also possible to prove this identity by algebraic manipulation from the formula for $\binom{n}{r}$ (see Exercise 23).

Remark: Pascal's identity, together with the initial conditions $\binom{n}{0} = \binom{n}{n} = 1$ for all integers n, can be used to recursively define binomial coefficients. This recursive definition is useful in the

$$
\begin{array}{c}
\binom{0}{0}\\[4pt]
\binom{1}{0}\ \binom{1}{1}\\[4pt]
\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}\\[4pt]
\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}\\[4pt]
\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4}\\[4pt]
\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5}\\[4pt]
\binom{6}{0}\ \binom{6}{1}\ \binom{6}{2}\ \binom{6}{3}\ \binom{6}{4}\ \binom{6}{5}\ \binom{6}{6}\\[4pt]
\binom{7}{0}\ \binom{7}{1}\ \binom{7}{2}\ \binom{7}{3}\ \binom{7}{4}\ \binom{7}{5}\ \binom{7}{6}\ \binom{7}{7}\\[4pt]
\binom{8}{0}\ \binom{8}{1}\ \binom{8}{2}\ \binom{8}{3}\ \binom{8}{4}\ \binom{8}{5}\ \binom{8}{6}\ \binom{8}{7}\ \binom{8}{8}\\[4pt]
\cdots
\end{array}
$$

(a)

By Pascal's identity:

$$\binom{6}{4} + \binom{6}{5} = \binom{7}{5}$$

$$
\begin{array}{c}
1\\
1 \quad 1\\
1 \quad 2 \quad 1\\
1 \quad 3 \quad 3 \quad 1\\
1 \quad 4 \quad 6 \quad 4 \quad 1\\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1\\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1\\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1\\
1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1\\
\cdots
\end{array}
$$

(b)

FIGURE 1 Pascal's triangle.

computation of binomial coefficients because only addition, and not multiplication, of integers is needed to use this recursive definition.

Pascal's identity is the basis for a geometric arrangement of the binomial coefficients in a triangle, as shown in Figure 1.

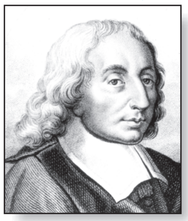
The nth row in the triangle consists of the binomial coefficients

$$\binom{n}{k},\ k = 0, 1, \ldots, n.$$

This triangle is known as **Pascal's triangle**, named after the French mathematician Blaise Pascal. Pascal's identity shows that when two adjacent binomial coefficients in this triangle are added, the binomial coefficient in the next row between these two coefficients is produced.

Pascal's triangle has a long and ancient history, predating Pascal by many centuries. In the East, binomial coefficients and Pascal's identity were known in the second century B.C.E. by the Indian mathematician Pingala. Later, Indian mathematicians included commentaries relating to Pascal's triangle in their books written in the first half of the last millennium. The Persian

Links



Source: National Library of Medicine

BLAISE PASCAL (1623–1662) Blaise Pascal was taught by his father, a tax collector in Rouen, France. He exhibited his talents at an early age, although his father, who had made discoveries in analytic geometry, kept mathematics books away from him to encourage other interests. At 16 Pascal discovered an important result concerning conic sections. At 18 he designed a calculating machine, which he built and sold. Pascal, along with Fermat, laid the foundations for the modern theory of probability. In this work, he made new discoveries concerning what is now called Pascal's triangle. In 1654, Pascal abandoned his mathematical pursuits to devote himself to theology. After this, he returned to mathematics only once. One night, distracted by a severe toothache, he sought comfort by studying the mathematical properties of the cycloid. Miraculously, his pain subsided, which he took as a sign of divine approval of the study of mathematics.

mathematician Al-Karaji and the multitalented Omar Khayyám wrote about Pascal's triangle in the eleventh and twelfth centuries, respectively; in Iran, Pascal's triangle is known as Khayyám's triangle. The triangle was known by the Chinese mathematician Jia Xian in the eleventh century and was written about in the 13th century by Yang Hui; in Chinese Pascal's triangle is often known as Yang Hui's triangle.

In the West, Pascal's triangle appears on the frontispiece of a 1527 book on business calculation written by the German scholar Petrus Apianus. In Italy, Pascal's triangle is called Tartaglia's triangle, after the Italian mathematician Niccolò Fontana Tartaglia who published the first few rows of the triangle in 1556. In his book *Traitè du triangle arithmétique*, published posthumously 1665, Pascal presented results about Pascal's triangle and used them to solve probability theory problems. Later French mathematicians named this triangle after Pascal; in 1730 Abraham de Moivre coined the name "Pascal's Arithmetic Triangle," which later became "Pascal's Triangle."

6.4.3 Other Identities Involving Binomial Coefficients

We conclude this section with combinatorial proofs of two of the many identities enjoyed by the binomial coefficients.

THEOREM 3 **VANDERMONDE'S IDENTITY** Let m, n, and r be nonnegative integers with r not exceeding either m or n. Then

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

Links

Remark: This identity was discovered by mathematician Alexandre-Théophile Vandermonde in the eighteenth century.

Proof: Suppose that there are m items in one set and n items in a second set. Then the total number of ways to pick r elements from the union of these sets is $\binom{m+n}{r}$.

Another way to pick r elements from the union is to pick k elements from the second set and then $r-k$ elements from the first set, where k is an integer with $0 \le k \le r$. Because there are $\binom{n}{k}$ ways to choose k elements from the second set and $\binom{m}{r-k}$ ways to choose $r-k$ elements from the first set, the product rule tells us that this can be done in $\binom{m}{r-k}\binom{n}{k}$ ways. Hence, the total number of ways to pick r elements from the union also equals $\sum_{k=0}^{r}\binom{m}{r-k}\binom{n}{k}$.

We have found two expressions for the number of ways to pick r elements from the union of a set with m items and a set with n items. Equating them gives us Vandermonde's identity. ◁

Corollary 4 follows from Vandermonde's identity.

ALEXANDRE-THÉOPHILE VANDERMONDE (1735–1796) Because Alexandre-Théophile Vandermonde was a sickly child, his physician father directed him to a career in music. However, he later developed an interest in mathematics. His complete mathematical work consists of four papers published in 1771–1772. These papers include fundamental contributions on the roots of equations, on the theory of determinants, and on the knight's tour problem (introduced in the exercises in Section 10.5). Vandermonde's interest in mathematics lasted for only 2 years. Afterward, he published papers on harmony, experiments with cold, and the manufacture of steel. He also became interested in politics, joining the cause of the French revolution and holding several different positions in government.

COROLLARY 4 If n is a nonnegative integer, then

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

Proof: We use Vandermonde's identity with $m = r = n$ to obtain

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

The last equality was obtained using the identity $\binom{n}{k} = \binom{n}{n-k}$. ◁

We can prove combinatorial identities by counting bit strings with different properties, as the proof of Theorem 4 will demonstrate.

THEOREM 4 Let n and r be nonnegative integers with $r \leq n$. Then

$$\binom{n+1}{r+1} = \sum_{j=r}^{n} \binom{j}{r}.$$

Proof: We use a combinatorial proof. By Example 14 in Section 6.3, the left-hand side, $\binom{n+1}{r+1}$, counts the bit strings of length $n + 1$ containing $r + 1$ ones.

We show that the right-hand side counts the same objects by considering the cases corresponding to the possible locations of the final 1 in a string with $r + 1$ ones. This final one must occur at position $r + 1, r + 2, \ldots$, or $n + 1$. Furthermore, if the last one is the kth bit there must be r ones among the first $k - 1$ positions. Consequently, by Example 14 in Section 6.3, there are $\binom{k-1}{r}$ such bit strings. Summing over k with $r + 1 \leq k \leq n + 1$, we find that there are

$$\sum_{k=r+1}^{n+1} \binom{k-1}{r} = \sum_{j=r}^{n} \binom{j}{r}$$

bit strings of length n containing exactly $r + 1$ ones. (Note that the last step follows from the change of variables $j = k - 1$.) Because the left-hand side and the right-hand side count the same objects, they are equal. This completes the proof. ◁

Exercises

1. Find the expansion of $(x + y)^4$
  a) using combinatorial reasoning, as in Example 1.
  b) using the binomial theorem.

2. Find the expansion of $(x + y)^5$
  a) using combinatorial reasoning, as in Example 1.
  b) using the binomial theorem.

3. Find the expansion of $(x + y)^6$.

4. Find the coefficient of x^5y^8 in $(x + y)^{13}$.

5. How many terms are there in the expansion of $(x + y)^{100}$ after like terms are collected?

6. What is the coefficient of x^7 in $(1 + x)^{11}$?

7. What is the coefficient of x^9 in $(2 - x)^{19}$?

8. What is the coefficient of x^8y^9 in the expansion of $(3x + 2y)^{17}$?

9. What is the coefficient of $x^{101}y^{99}$ in the expansion of $(2x - 3y)^{200}$?

10. Use the binomial theorem to expand $(3x - y^2)^4$ into a sum of terms of the form $cx^a y^b$, where c is a real number and a and b are nonnegative integers.

11. Use the binomial theorem to expand $(3x^4 - 2y^3)^5$ into a sum of terms of the form $cx^a y^b$, where c is a real number and a and b are nonnegative integers.

12. Use the binomial theorem to find the coefficient of $x^a y^b$ in the expansion of $(5x^2 + 2y^3)^6$, where

a) $a = 6, b = 9$.
b) $a = 2, b = 15$.
c) $a = 3, b = 12$.
d) $a = 12, b = 0$.
e) $a = 8, b = 9$.

13. Use the binomial theorem to find the coefficient of $x^a y^b$ in the expansion of $(2x^3 - 4y^2)^7$, where

a) $a = 9, b = 8$.
b) $a = 8, b = 9$.
c) $a = 0, b = 14$.
d) $a = 12, b = 6$.
e) $a = 18, b = 2$.

***14.** Give a formula for the coefficient of x^k in the expansion of $(x + 1/x)^{100}$, where k is an integer.

***15.** Give a formula for the coefficient of x^k in the expansion of $(x^2 - 1/x)^{100}$, where k is an integer.

16. The row of Pascal's triangle containing the binomial coefficients $\binom{10}{k}$, $0 \le k \le 10$, is:

1 10 45 120 210 252 210 120 45 10 1

Use Pascal's identity to produce the row immediately following this row in Pascal's triangle.

17. What is the row of Pascal's triangle containing the binomial coefficients $\binom{9}{k}$, $0 \le k \le 9$?

18. Show that if n is a positive integer, then $1 = \binom{n}{0} < \binom{n}{1} < \cdots < \binom{n}{\lfloor n/2 \rfloor} = \binom{n}{\lceil n/2 \rceil} > \cdots > \binom{n}{n-1} > \binom{n}{n} = 1$.

19. Show that $\binom{n}{k} \le 2^n$ for all positive integers n and all integers k with $0 \le k \le n$.

20. a) Use Exercise 18 and Corollary 1 to show that if n is an integer greater than 1, then $\binom{n}{\lfloor n/2 \rfloor} \ge 2^n/n$.

b) Conclude from part (a) that if n is a positive integer, then $\binom{2n}{n} \ge 4^n/2n$.

21. Show that if n and k are integers with $1 \le k \le n$, then $\binom{n}{k} \le n^k/2^{k-1}$.

22. Suppose that b is an integer with $b \ge 7$. Use the binomial theorem and the appropriate row of Pascal's triangle to find the base-b expansion of $(11)_b^4$ [that is, the fourth power of the number $(11)_b$ in base-b notation].

23. Prove Pascal's identity, using the formula for $\binom{n}{r}$.

24. Suppose that k and n are integers with $1 \le k < n$. Prove the **hexagon identity**

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

which relates terms in Pascal's triangle that form a hexagon.

25. Prove that if n and k are integers with $1 \le k \le n$, then $k\binom{n}{k} = n\binom{n-1}{k-1}$,

a) using a combinatorial proof. [*Hint:* Show that the two sides of the identity count the number of ways to select a subset with k elements from a set with n elements and then an element of this subset.]

b) using an algebraic proof based on the formula for $\binom{n}{r}$ given in Theorem 2 in Section 6.3.

26. Prove the identity $\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$, whenever n, r, and k are nonnegative integers with $r \le n$ and $k \le r$,

a) using a combinatorial argument.

b) using an argument based on the formula for the number of r-combinations of a set with n elements.

27. Show that if n and k are positive integers, then

$$\binom{n+1}{k} = (n+1)\binom{n}{k-1} \Big/ k.$$

Use this identity to construct an inductive definition of the binomial coefficients.

28. Show that if p is a prime and k is an integer such that $1 \le k \le p - 1$, then p divides $\binom{p}{k}$.

29. Let n be a positive integer. Show that

$$\binom{2n}{n+1} + \binom{2n}{n} = \binom{2n+2}{n+1}\Big/2.$$

***30.** Let n and k be integers with $1 \le k \le n$. Show that

$$\sum_{k=1}^{n} \binom{n}{k}\binom{n}{k-1} = \binom{2n+2}{n+1}\Big/2 - \binom{2n}{n}.$$

***31.** Prove the **hockeystick identity**

$$\sum_{k=0}^{r} \binom{n+k}{k} = \binom{n+r+1}{r}$$

whenever n and r are positive integers,

a) using a combinatorial argument.

b) using Pascal's identity.

32. Show that if n is a positive integer, then $\binom{2n}{2} = 2\binom{n}{2} + n^2$

a) using a combinatorial argument.

b) by algebraic manipulation.

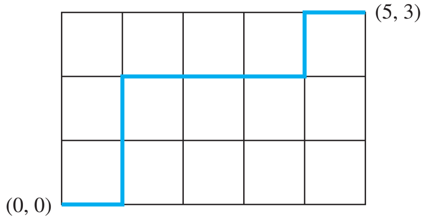
***33.** Give a combinatorial proof that $\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$. [*Hint:* Count in two ways the number of ways to select a committee and to then select a leader of the committee.]

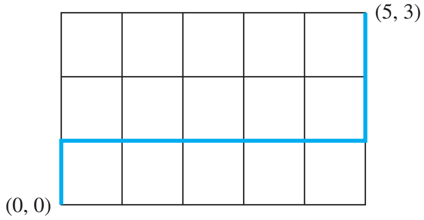
***34.** Give a combinatorial proof that $\sum_{k=1}^{n} k\binom{n}{k}^2 = n\binom{2n-1}{n-1}$. [*Hint:* Count in two ways the number of ways to select a committee, with n members from a group of n mathematics professors and n computer science professors, such that the chairperson of the committee is a mathematics professor.]

35. Show that a nonempty set has the same number of subsets with an odd number of elements as it does subsets with an even number of elements.

***36.** Prove the binomial theorem using mathematical induction.

37. In this exercise we will count the number of paths in the *xy* plane between the origin (0, 0) and point (m, n), where m and n are nonnegative integers, such that each path is made up of a series of steps, where each step is a move one unit to the right or a move one unit upward. (No moves to the left or downward are allowed.) Two such paths from (0, 0) to (5, 3) are illustrated here.





a) Show that each path of the type described can be represented by a bit string consisting of m 0s and n 1s, where a 0 represents a move one unit to the right and a 1 represents a move one unit upward.

b) Conclude from part (a) that there are $\binom{m+n}{n}$ paths of the desired type.

38. Use Exercise 37 to give an alternative proof of Corollary 2 in Section 6.3, which states that $\binom{n}{k} = \binom{n}{n-k}$ whenever k is an integer with $0 \le k \le n$. [*Hint:* Consider the number of paths of the type described in Exercise 37 from (0, 0) to $(n-k, k)$ and from (0, 0) to $(k, n-k)$.]

39. Use Exercise 37 to prove Theorem 4. [*Hint:* Count the number of paths with n steps of the type described in Exercise 37. Every such path must end at one of the points $(n-k, k)$ for $k = 0, 1, 2, \ldots, n$.]

40. Use Exercise 37 to prove Pascal's identity. [*Hint:* Show that a path of the type described in Exercise 37 from (0, 0) to $(n+1-k, k)$ passes through either $(n+1-k, k-1)$ or $(n-k, k)$, but not through both.]

41. Use Exercise 37 to prove the hockeystick identity from Exercise 31. [*Hint:* First, note that the number of paths from (0, 0) to $(n+1, r)$ equals $\binom{n+1+r}{r}$. Second, count the number of paths by summing the number of these paths that start by going k units upward for $k = 0, 1, 2, \ldots, r$.]

42. Give a combinatorial proof that if n is a positive integer then $\sum_{k=0}^{n} k^2\binom{n}{k} = n(n+1)2^{n-2}$. [*Hint:* Show that both sides count the ways to select a subset of a set of n elements together with two not necessarily distinct elements from this subset. Furthermore, express the right-hand side as $n(n-1)2^{n-2} + n2^{n-1}$.]

***43.** Determine a formula involving binomial coefficients for the nth term of a sequence if its initial terms are those listed. [*Hint:* Looking at Pascal's triangle will be helpful. Although infinitely many sequences start with a specified set of terms, each of the following lists is the start of a sequence of the type desired.]

a) 1, 3, 6, 10, 15, 21, 28, 36, 45, 55, 66, …

b) 1, 4, 10, 20, 35, 56, 84, 120, 165, 220, …

c) 1, 2, 6, 20, 70, 252, 924, 3432, 12870, 48620, …

d) 1, 1, 2, 3, 6, 10, 20, 35, 70, 126, …

e) 1, 1, 1, 3, 1, 5, 15, 35, 1, 9, …

f) 1, 3, 15, 84, 495, 3003, 18564, 116280, 735471, 4686825, …

6.5 Generalized Permutations and Combinations

6.5.1 Introduction

Links

In many counting problems, elements may be used repeatedly. For instance, a letter or digit may be used more than once on a license plate. When a dozen donuts are selected, each variety can be chosen repeatedly. This contrasts with the counting problems discussed earlier in the chapter where we considered only permutations and combinations in which each item could be used at most once. In this section we will show how to solve counting problems where elements may be used more than once.

Also, some counting problems involve indistinguishable elements. For instance, to count the number of ways the letters of the word *SUCCESS* can be rearranged, the placement of identical letters must be considered. This contrasts with the counting problems discussed earlier where all elements were considered distinguishable. In this section we will describe how to solve counting problems in which some elements are indistinguishable.

Moreover, in this section we will explain how to solve another important class of counting problems, problems involving counting the ways distinguishable elements can be placed in boxes. An example of this type of problem is the number of different ways poker hands can be dealt to four players.

Taken together, the methods described earlier in this chapter and the methods introduced in this section form a useful toolbox for solving a wide range of counting problems. When the additional methods discussed in Chapter 8 are added to this arsenal, you will be able to solve a large percentage of the counting problems that arise in a wide range of areas of study.

6.5.2 Permutations with Repetition

Counting permutations when repetition of elements is allowed can easily be done using the product rule, as Example 1 shows.

EXAMPLE 1 How many strings of length r can be formed from the uppercase letters of the English alphabet?

Solution: By the product rule, because there are 26 uppercase English letters, and because each letter can be used repeatedly, we see that there are 26^r strings of uppercase English letters of length r. ◀

The number of r-permutations of a set with n elements when repetition is allowed is given in Theorem 1.

THEOREM 1 The number of r-permutations of a set of n objects with repetition allowed is n^r.

Proof: There are n ways to select an element of the set for each of the r positions in the r-permutation when repetition is allowed, because for each choice all n objects are available. Hence, by the product rule there are n^r r-permutations when repetition is allowed. ◁

6.5.3 Combinations with Repetition

Consider these examples of combinations with repetition of elements allowed.

EXAMPLE 2 How many ways are there to select four pieces of fruit from a bowl containing apples, oranges, and pears if the order in which the pieces are selected does not matter, only the type of fruit and not the individual piece matters, and there are at least four pieces of each type of fruit in the bowl?

Solution: To solve this problem we list all the ways possible to select the fruit. There are 15 ways:

4 apples	4 oranges	4 pears
3 apples, 1 orange	3 apples, 1 pear	3 oranges, 1 apple
3 oranges, 1 pear	3 pears, 1 apple	3 pears, 1 orange
2 apples, 2 oranges	2 apples, 2 pears	2 oranges, 2 pears
2 apples, 1 orange, 1 pear	2 oranges, 1 apple, 1 pear	2 pears, 1 apple, 1 orange

The solution is the number of 4-combinations with repetition allowed from a three-element set, $\{apple, orange, pear\}$. ◀

To solve more complex counting problems of this type, we need a general method for counting the r-combinations of an n-element set. In Example 3 we will illustrate such a method.

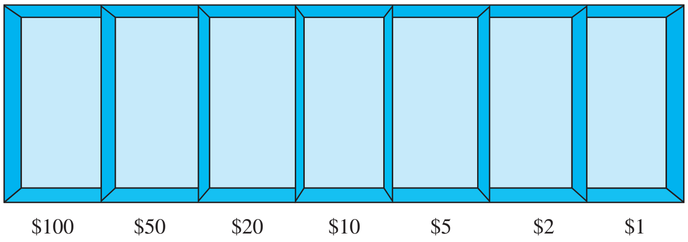


FIGURE 1 **Cash box with seven types of bills.**

EXAMPLE 3 How many ways are there to select five bills from a cash box containing $1 bills, $2 bills, $5 bills, $10 bills, $20 bills, $50 bills, and $100 bills? Assume that the order in which the bills are chosen does not matter, that the bills of each denomination are indistinguishable, and that there are at least five bills of each type.

Solution: Because the order in which the bills are selected does not matter and seven different types of bills can be selected as many as five times, this problem involves counting 5-combinations with repetition allowed from a set with seven elements. Listing all possibilities would be tedious, because there are a large number of solutions. Instead, we will illustrate the use of a technique for counting combinations with repetition allowed.

Suppose that a cash box has seven compartments, one to hold each type of bill, as illustrated in Figure 1. These compartments are separated by six dividers, as shown in the picture. The choice of five bills corresponds to placing five markers in the compartments holding different types of bills. Figure 2 illustrates this correspondence for three different ways to select five bills, where the six dividers are represented by bars and the five bills by stars.

The number of ways to select five bills corresponds to the number of ways to arrange six bars and five stars in a row with a total of 11 positions. Consequently, the number of ways to select the five bills is the number of ways to select the positions of the five stars from the 11

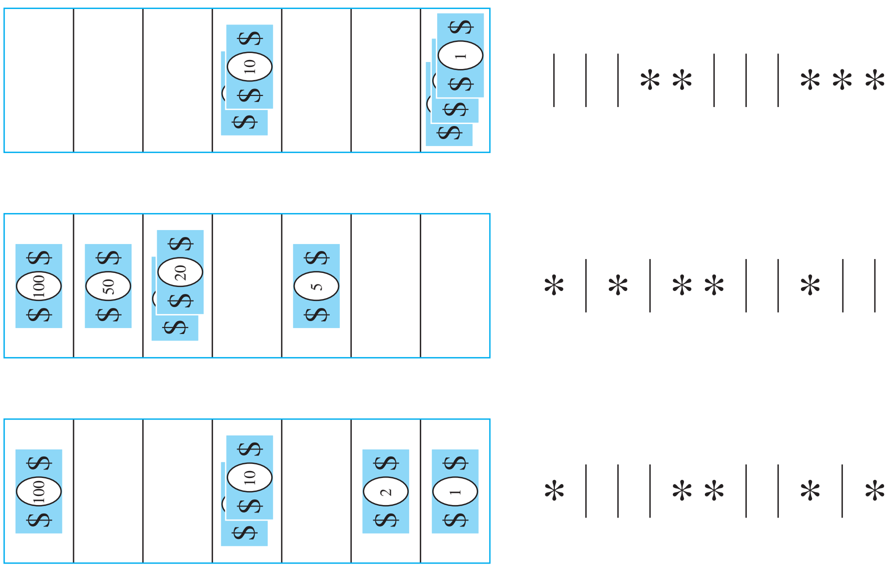


FIGURE 2 **Examples of ways to select five bills.**

positions. This corresponds to the number of unordered selections of 5 objects from a set of 11 objects, which can be done in $C(11, 5)$ ways. Consequently, there are

$$C(11, 5) = \frac{11!}{5!\,6!} = 462$$

ways to choose five bills from the cash box with seven types of bills. ◀

Theorem 2 generalizes this discussion.

THEOREM 2 There are $C(n + r - 1, r) = C(n + r - 1, n - 1)$ r-combinations from a set with n elements when repetition of elements is allowed.

Proof: Each r-combination of a set with n elements when repetition is allowed can be represented by a list of $n - 1$ bars and r stars. The $n - 1$ bars are used to mark off n different cells, with the ith cell containing a star for each time the ith element of the set occurs in the combination. For instance, a 6-combination of a set with four elements is represented with three bars and six stars. Here

$$**|*|\,|***$$

represents the combination containing exactly two of the first element, one of the second element, none of the third element, and three of the fourth element of the set.

As we have seen, each different list containing $n - 1$ bars and r stars corresponds to an r-combination of the set with n elements, when repetition is allowed. The number of such lists is $C(n - 1 + r, r)$, because each list corresponds to a choice of the r positions to place the r stars from the $n - 1 + r$ positions that contain r stars and $n - 1$ bars. The number of such lists is also equal to $C(n - 1 + r, n - 1)$, because each list corresponds to a choice of the $n - 1$ positions to place the $n - 1$ bars. ◁

Examples 4–6 show how Theorem 2 is applied.

EXAMPLE 4 Suppose that a cookie shop has four different kinds of cookies. How many different ways can six cookies be chosen? Assume that only the type of cookie, and not the individual cookies or the order in which they are chosen, matters.

Extra Examples

Solution: The number of ways to choose six cookies is the number of 6-combinations of a set with four elements. From Theorem 2 this equals $C(4 + 6 - 1, 6) = C(9, 6)$. Because

$$C(9, 6) = C(9, 3) = \frac{9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3} = 84,$$

there are 84 different ways to choose the six cookies. ◀

Theorem 2 can also be used to find the number of solutions of certain linear equations where the variables are integers subject to constraints. This is illustrated by Example 5.

EXAMPLE 5 How many solutions does the equation

$$x_1 + x_2 + x_3 = 11$$

have, where x_1, x_2, and x_3 are nonnegative integers?

Solution: To count the number of solutions, we note that a solution corresponds to a way of selecting 11 items from a set with three elements so that x_1 items of type one, x_2 items of type two, and x_3 items of type three are chosen. Hence, the number of solutions is equal to the number of 11-combinations with repetition allowed from a set with three elements. From Theorem 2 it follows that there are

$$C(3 + 11 - 1, 11) = C(13, 11) = C(13, 2) = \frac{13 \cdot 12}{1 \cdot 2} = 78$$

solutions.

The number of solutions of this equation can also be found when the variables are subject to constraints. For instance, we can find the number of solutions where the variables are integers with $x_1 \geq 1$, $x_2 \geq 2$, and $x_3 \geq 3$. A solution to the equation subject to these constraints corresponds to a selection of 11 items with x_1 items of type one, x_2 items of type two, and x_3 items of type three, where, in addition, there is at least one item of type one, two items of type two, and three items of type three. So, a solution corresponds to a choice of one item of type one, two of type two, and three of type three, together with a choice of five additional items of any type. By Theorem 2 this can be done in

$$C(3 + 5 - 1, 5) = C(7, 5) = C(7, 2) = \frac{7 \cdot 6}{1 \cdot 2} = 21$$

ways. Thus, there are 21 solutions of the equation subject to the given constraints. ◀

Example 6 shows how counting the number of combinations with repetition allowed arises in determining the value of a variable that is incremented each time a certain type of nested loop is traversed.

EXAMPLE 6 What is the value of k after the following pseudocode has been executed?

```
k := 0
for i_1 := 1 to n
   for i_2 := 1 to i_1
       .
       .
       .
      for i_m := 1 to i_{m-1}
         k := k + 1
```

Solution: Note that the initial value of k is 0 and that 1 is added to k each time the nested loop is traversed with a sequence of integers $i_1, i_2, \ldots, i_m$ such that

$$1 \leq i_m \leq i_{m-1} \leq \cdots \leq i_1 \leq n.$$

The number of such sequences of integers is the number of ways to choose m integers from $\{1, 2, \ldots, n\}$, with repetition allowed. (To see this, note that once such a sequence has been selected, if we order the integers in the sequence in nondecreasing order, this uniquely defines an assignment of $i_m, i_{m-1}, \ldots, i_1$. Conversely, every such assignment corresponds to a unique unordered set.) Hence, from Theorem 2, it follows that $k = C(n + m - 1, m)$ after this code has been executed. ◀

The formulae for the numbers of ordered and unordered selections of r elements, chosen with and without repetition allowed from a set with n elements, are shown in Table 1.

TABLE 1 Combinations and Permutations With and Without Repetition.

Type	*Repetition Allowed?*	*Formula*
r-permutations	No	$\frac{n!}{(n-r)!}$
r-combinations	No	$\frac{n!}{r!\,(n-r)!}$
r-permutations	Yes	n^r
r-combinations	Yes	$\frac{(n+r-1)!}{r!\,(n-1)!}$

6.5.4 Permutations with Indistinguishable Objects

Some elements may be indistinguishable in counting problems. When this is the case, care must be taken to avoid counting things more than once. Consider Example 7.

EXAMPLE 7 How many different strings can be made by reordering the letters of the word *SUCCESS*?

Extra Examples

Solution: Because some of the letters of *SUCCESS* are the same, the answer is *not* given by the number of permutations of seven letters. This word contains three Ss, two Cs, one U, and one E. To determine the number of different strings that can be made by reordering the letters, first note that the three Ss can be placed among the seven positions in $C(7, 3)$ different ways, leaving four positions free. Then the two Cs can be placed in $C(4, 2)$ ways, leaving two free positions. The U can be placed in $C(2, 1)$ ways, leaving just one position free. Hence E can be placed in $C(1, 1)$ way. Consequently, from the product rule, the number of different strings that can be made is

$$\begin{aligned} C(7,3)C(4,2)C(2,1)C(1,1) &= \frac{7!}{3!\,4!} \cdot \frac{4!}{2!\,2!} \cdot \frac{2!}{1!\,1!} \cdot \frac{1!}{1!\,0!} \\ &= \frac{7!}{3!\,2!\,1!\,1!} \\ &= 420. \end{aligned}$$

◄

We can prove Theorem 3 using the same sort of reasoning as in Example 7.

THEOREM 3 The number of different permutations of n objects, where there are n_1 indistinguishable objects of type 1, n_2 indistinguishable objects of type 2, ..., and n_k indistinguishable objects of type k, is

$$\frac{n!}{n_1!\,n_2!\cdots n_k!}.$$

Proof: To determine the number of permutations, first note that the n_1 objects of type one can be placed among the n positions in $C(n, n_1)$ ways, leaving $n - n_1$ positions free. Then the objects of type two can be placed in $C(n - n_1, n_2)$ ways, leaving $n - n_1 - n_2$ positions free. Continue placing the objects of type three, ..., type $k - 1$, until at the last stage, n_k objects of type k can be placed in $C(n - n_1 - n_2 - \cdots - n_{k-1}, n_k)$ ways. Hence, by the product rule, the total number of different permutations is

$$\begin{aligned} & C(n, n_1)C(n - n_1, n_2) \cdots C(n - n_1 - \cdots - n_{k-1}, n_k) \\ & \quad = \frac{n!}{n_1!\,(n - n_1)!} \frac{(n - n_1)!}{n_2!\,(n - n_1 - n_2)!} \cdots \frac{(n - n_1 - \cdots - n_{k-1})!}{n_k!\,0!} \\ & \quad = \frac{n!}{n_1!\,n_2! \cdots n_k!}. \end{aligned}$$

◁

6.5.5 Distributing Objects into Boxes

Links

Many counting problems can be solved by enumerating the ways objects can be placed into boxes (where the order these objects are placed into the boxes does not matter). The objects can be either *distinguishable*, that is, different from each other, or *indistinguishable*, that is, considered identical. Distinguishable objects are sometimes said to be *labeled*, whereas indistinguishable objects are said to be *unlabeled*. Similarly, boxes can be *distinguishable*, that is, different, or *indistinguishable*, that is, identical. Distinguishable boxes are often said to be *labeled*, while indistinguishable boxes are said to be *unlabeled*. When you solve a counting problem using the model of distributing objects into boxes, you need to determine whether the objects are distinguishable and whether the boxes are distinguishable. Although the context of the counting problem makes these two decisions clear, counting problems are sometimes ambiguous and it may be unclear which model applies. In such a case it is best to state whatever assumptions you are making and explain why the particular model you choose conforms to your assumptions.

Extra Examples

We will see that there are closed formulae for counting the ways to distribute objects, distinguishable or indistinguishable, into distinguishable boxes. We are not so lucky when we count the ways to distribute objects, distinguishable or indistinguishable, into indistinguishable boxes; there are no closed formulae to use in these cases.

Remark: A **closed formula** is an expression that can be evaluated using a finite number of operations and that includes numbers, variables, and values of functions, where the operations and functions belong to a generally accepted set that can depend on the context. In this book, we include the usual arithmetic operations, rational powers, exponential and logarithmic functions, trigonometric functions, and the factorial function. We do not allow infinite series to be included in closed formulae.

DISTINGUISHABLE OBJECTS AND DISTINGUISHABLE BOXES We first consider the case when distinguishable objects are placed into distinguishable boxes. Consider Example 8 in which the objects are cards and the boxes are hands of players.

EXAMPLE 8 How many ways are there to distribute hands of 5 cards to each of four players from the standard deck of 52 cards?

Solution: We will use the product rule to solve this problem. To begin, note that the first player can be dealt 5 cards in $C(52, 5)$ ways. The second player can be dealt 5 cards in $C(47, 5)$ ways, because only 47 cards are left. The third player can be dealt 5 cards in $C(42, 5)$ ways. Finally, the fourth player can be dealt 5 cards in $C(37, 5)$ ways. Hence, the total number of ways to deal four players 5 cards each is

$$\begin{aligned} C(52, 5)C(47, 5)C(42, 5)C(37, 5) &= \frac{52!}{47!\,5!} \cdot \frac{47!}{42!\,5!} \cdot \frac{42!}{37!\,5!} \cdot \frac{37!}{32!\,5!} \\ &= \frac{52!}{5!\,5!\,5!\,5!\,32!}. \end{aligned}$$

◀

Remark: The solution to Example 8 equals the number of permutations of 52 objects, with 5 indistinguishable objects of each of four different types, and 32 objects of a fifth type. This equality can be seen by defining a one-to-one correspondence between permutations of this type and distributions of cards to the players. To define this correspondence, first order the cards from 1 to 52. Then cards dealt to the first player correspond to the cards in the positions assigned to objects of the first type in the permutation. Similarly, cards dealt to the second, third, and fourth players, respectively, correspond to cards in the positions assigned to objects of the second, third, and fourth type, respectively. The cards not dealt to any player correspond to cards in the positions assigned to objects of the fifth type. The reader should verify that this is a one-to-one correspondence.

Example 8 is a typical problem that involves distributing distinguishable objects into distinguishable boxes. The distinguishable objects are the 52 cards, and the five distinguishable boxes are the hands of the four players and the rest of the deck. Counting problems that involve distributing distinguishable objects into boxes can be solved using Theorem 4.

THEOREM 4 The number of ways to distribute n distinguishable objects into k distinguishable boxes so that n_i objects are placed into box i, $i = 1, 2, \ldots, k$, equals

$$\frac{n!}{n_1!\,n_2!\cdots n_k!}.$$

Theorem 4 can be proved using the product rule. We leave the details as Exercise 49. It can also be proved (see Exercise 50) by setting up a one-to-one correspondence between the permutations counted by Theorem 3 and the ways to distribute objects counted by Theorem 4.

INDISTINGUISHABLE OBJECTS AND DISTINGUISHABLE BOXES Counting the number of ways of placing n indistinguishable objects into k distinguishable boxes turns out to be the same as counting the number of n-combinations for a set with k elements when repetitions are allowed. The reason behind this is that there is a one-to-one correspondence between n-combinations from a set with k elements when repetition is allowed and the ways to place n indistinguishable balls into k distinguishable boxes. To set up this correspondence, we put a ball in the ith bin each time the ith element of the set is included in the n-combination.

EXAMPLE 9 How many ways are there to place 10 indistinguishable balls into eight distinguishable bins?

Solution: The number of ways to place 10 indistinguishable balls into eight bins equals the number of 10-combinations from a set with eight elements when repetition is allowed. Consequently, there are

$$C(8 + 10 - 1, 10) = C(17, 10) = \frac{17!}{10!7!} = 19{,}448.$$

◀

This means that there are $C(n + r - 1, n - 1)$ ways to place r indistinguishable objects into n distinguishable boxes.

DISTINGUISHABLE OBJECTS AND INDISTINGUISHABLE BOXES Counting the ways to place n distinguishable objects into k indistinguishable boxes is more difficult than counting the ways to place objects, distinguishable or indistinguishable objects, into distinguishable boxes. We illustrate this with an example.

EXAMPLE 10 How many ways are there to put four different employees into three indistinguishable offices, when each office can contain any number of employees?

Solution: We will solve this problem by enumerating all the ways these employees can be placed into the offices. We represent the four employees by A, B, C, and D. First, we note that we can distribute employees so that all four are put into one office, three are put into one office and a fourth is put into a second office, two employees are put into one office and two put into a second office, and finally, two are put into one office, and one each put into the other two offices. Each way to distribute these employees to these offices can be represented by a way to partition the elements A, B, C, and D into disjoint subsets.

We can put all four employees into one office in exactly one way, represented by $\{\{A, B, C, D\}\}$. We can put three employees into one office and the fourth employee into a different office in exactly four ways, represented by $\{\{A, B, C\}, \{D\}\}$, $\{\{A, B, D\}, \{C\}\}$, $\{\{A, C, D\}, \{B\}\}$, and $\{\{B, C, D\}, \{A\}\}$. We can put two employees into one office and two into a second office in exactly three ways, represented by $\{\{A, B\}, \{C, D\}\}$, $\{\{A, C\}, \{B, D\}\}$, and $\{\{A, D\}, \{B, C\}\}$. Finally, we can put two employees into one office, and one each into each of the remaining two offices in six ways, represented by $\{\{A, B\}, \{C\}, \{D\}\}$, $\{\{A, C\}, \{B\}, \{D\}\}$, $\{\{A, D\}, \{B\}, \{C\}\}$, $\{\{B, C\}, \{A\}, \{D\}\}$, $\{\{B, D\}\}, \{A\}, \{C\}\}$, and $\{\{C, D\}, \{A\}, \{B\}\}$.

Counting all the possibilities, we find that there are 14 ways to put four different employees into three indistinguishable offices. Another way to look at this problem is to look at the number of offices into which we put employees. Note that there are six ways to put four different employees into three indistinguishable offices so that no office is empty, seven ways to put four different employees into two indistinguishable offices so that no office is empty, and one way to put four employees into one office so that it is not empty. ◀

There is no simple closed formula for the number of ways to distribute n distinguishable objects into j indistinguishable boxes. However, there is a formula involving a summation, which we will now describe. Let $S(n, j)$ denote the number of ways to distribute n distinguishable objects into j indistinguishable boxes so that no box is empty. The numbers $S(n, j)$ are called **Stirling numbers of the second kind**. For instance, Example 10 shows that $S(4, 3) = 6$, $S(4, 2) = 7$, and $S(4, 1) = 1$. We see that the number of ways to distribute n distinguishable objects into k indistinguishable boxes (where the number of boxes that are nonempty equals $k, k - 1, \ldots, 2$, or 1) equals $\sum_{j=1}^{k} S(n, j)$. For instance, following the reasoning in Example 10, the number of ways to distribute four distinguishable objects into three indistinguishable boxes equals $S(4, 1) + S(4, 2) + S(4, 3) = 1 + 7 + 6 = 14$. Using the inclusion–exclusion principle (see Section 8.6) it can be shown that

$$S(n, j) = \frac{1}{j!} \sum_{i=0}^{j-1} (-1)^i \binom{j}{i} (j - i)^n.$$

Consequently, the number of ways to distribute n distinguishable objects into k indistinguishable boxes equals

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

Remark: The reader may be curious about the Stirling numbers of the first kind. A combinatorial definition of the **signless Stirling numbers of the first kind**, the absolute values of the Stirling numbers of the first kind, can be found in the preamble to Exercise 47 in the Supplementary Exercises. For the definition of Stirling numbers of the first kind, for more information about Stirling numbers of the second kind, and to learn more about Stirling numbers of the first kind and the relationship between Stirling numbers of the first and second kind, see combinatorics textbooks such as [Bó07], [Br99], and [RoTe05], and Chapter 6 in [MiRo91].

INDISTINGUISHABLE OBJECTS AND INDISTINGUISHABLE BOXES Some counting problems can be solved by determining the number of ways to distribute indistinguishable objects into indistinguishable boxes. We illustrate this principle with an example.

EXAMPLE 11 How many ways are there to pack six copies of the same book into four identical boxes, where a box can contain as many as six books?

Solution: We will enumerate all ways to pack the books. For each way to pack the books, we will list the number of books in the box with the largest number of books, followed by the numbers of books in each box containing at least one book, in order of decreasing number of books in a box. The ways we can pack the books are

6	3, 2, 1
5, 1	3, 1, 1, 1
4, 2	2, 2, 2
4, 1, 1	2, 2, 1, 1.
3, 3	

For example, 4, 1, 1 indicates that one box contains four books, a second box contains a single book, and a third box contains a single book (and the fourth box is empty). We conclude that there are nine allowable ways to pack the books, because we have listed them all. ◀

Observe that distributing n indistinguishable objects into k indistinguishable boxes is the same as writing n as the sum of at most k positive integers in nonincreasing order. If $a_1 + a_2 + \cdots + a_j = n$, where $a_1, a_2, \ldots, a_j$ are positive integers with $a_1 \geq a_2 \geq \cdots \geq a_j$, we say that $a_1, a_2, \ldots, a_j$ is a **partition** of the positive integer n into j positive integers. We see that if $p_k(n)$ is the number of partitions of n into at most k positive integers, then there are $p_k(n)$ ways to distribute n indistinguishable objects into k indistinguishable boxes. No simple closed formula exists for this number. For more information about partitions of positive integers, see [Ro11].

Exercises

1. In how many different ways can five elements be selected in order from a set with three elements when repetition is allowed?

2. In how many different ways can five elements be selected in order from a set with five elements when repetition is allowed?

3. How many strings of six letters are there?

4. Every day a student randomly chooses a sandwich for lunch from a pile of wrapped sandwiches. If there are six kinds of sandwiches, how many different ways are there for the student to choose sandwiches for the seven days of a week if the order in which the sandwiches are chosen matters?

5. How many ways are there to assign three jobs to five employees if each employee can be given more than one job?

6. How many ways are there to select five unordered elements from a set with three elements when repetition is allowed?

7. How many ways are there to select three unordered elements from a set with five elements when repetition is allowed?

8. How many different ways are there to choose a dozen donuts from the 21 varieties at a donut shop?

9. A bagel shop has onion bagels, poppy seed bagels, egg bagels, salty bagels, pumpernickel bagels, sesame seed bagels, raisin bagels, and plain bagels. How many ways are there to choose

a) six bagels?
b) a dozen bagels?
c) two dozen bagels?
d) a dozen bagels with at least one of each kind?
e) a dozen bagels with at least three egg bagels and no more than two salty bagels?

10. A croissant shop has plain croissants, cherry croissants, chocolate croissants, almond croissants, apple croissants, and broccoli croissants. How many ways are there to choose

a) a dozen croissants?

b) three dozen croissants?

c) two dozen croissants with at least two of each kind?

d) two dozen croissants with no more than two broccoli croissants?

e) two dozen croissants with at least five chocolate croissants and at least three almond croissants?

f) two dozen croissants with at least one plain croissant, at least two cherry croissants, at least three chocolate croissants, at least one almond croissant, at least two apple croissants, and no more than three broccoli croissants?

11. How many ways are there to choose eight coins from a piggy bank containing 100 identical pennies and 80 identical nickels?

12. How many different combinations of pennies, nickels, dimes, quarters, and half dollars can a piggy bank contain if it has 20 coins in it?

13. A book publisher has 3000 copies of a discrete mathematics book. How many ways are there to store these books in their three warehouses if the copies of the book are indistinguishable?

14. How many solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 = 17,$$

where x_1, x_2, x_3, and x_4 are nonnegative integers?

15. How many solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 = 21,$$

where x_i, $i = 1, 2, 3, 4, 5$, is a nonnegative integer such that

a) $x_1 \geq 1$?

b) $x_i \geq 2$ for $i = 1, 2, 3, 4, 5$?

c) $0 \leq x_1 \leq 10$?

d) $0 \leq x_1 \leq 3$, $1 \leq x_2 < 4$, and $x_3 \geq 15$?

16. How many solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 29,$$

where x_i, $i = 1, 2, 3, 4, 5, 6$, is a nonnegative integer such that

a) $x_i > 1$ for $i = 1, 2, 3, 4, 5, 6$?

b) $x_1 \geq 1$, $x_2 \geq 2$, $x_3 \geq 3$, $x_4 \geq 4$, $x_5 > 5$, and $x_6 \geq 6$?

c) $x_1 \leq 5$?

d) $x_1 < 8$ and $x_2 > 8$?

17. How many strings of 10 ternary digits (0, 1, or 2) are there that contain exactly two 0s, three 1s, and five 2s?

18. How many strings of 20-decimal digits are there that contain two 0s, four 1s, three 2s, one 3, two 4s, three 5s, two 7s, and three 9s?

19. Suppose that a large family has 14 children, including two sets of identical triplets, three sets of identical twins, and two individual children. How many ways are there to seat these children in a row of chairs if the identical triplets or twins cannot be distinguished from one another?

20. How many solutions are there to the inequality

$$x_1 + x_2 + x_3 \leq 11,$$

where x_1, x_2, and x_3 are nonnegative integers? [*Hint:* Introduce an auxiliary variable x_4 such that $x_1 + x_2 + x_3 + x_4 = 11$.]

21. A Swedish tour guide has devised a clever way for his clients to recognize him. He owns 13 pairs of shoes of the same style, customized so that each pair has a unique color. How many ways are there for him to choose a left shoe and a right shoe from these 13 pairs

a) without restrictions and which color is on which foot matters?

b) so that the colors of the left and right shoe are different and which color is on which foot matters?

c) so that the colors of the left and right shoe are different but which color is on which foot does not matter?

d) without restrictions, but which color is on which foot does not matter?

∗22. In how many ways can an airplane pilot be scheduled for five days of flying in October if he cannot work on consecutive days?

23. How many ways are there to distribute six indistinguishable balls into nine distinguishable bins?

24. How many ways are there to distribute 12 indistinguishable balls into six distinguishable bins?

25. How many ways are there to distribute 12 distinguishable objects into six distinguishable boxes so that two objects are placed in each box?

26. How many ways are there to distribute 15 distinguishable objects into five distinguishable boxes so that the boxes have one, two, three, four, and five objects in them, respectively.

27. How many positive integers less than 1,000,000 have the sum of their digits equal to 19?

28. How many positive integers less than 1,000,000 have exactly one digit equal to 9 and have a sum of digits equal to 13?

29. There are 10 questions on a discrete mathematics final exam. How many ways are there to assign scores to the problems if the sum of the scores is 100 and each question is worth at least 5 points?

30. Show that there are $C(n + r - q_1 - q_2 - \cdots - q_r - 1, n - q_1 - q_2 - \cdots - q_r)$ different unordered selections of n objects of r different types that include at least q_1 objects of type one, q_2 objects of type two, …, and q_r objects of type r.

31. How many different bit strings can be transmitted if the string must begin with a 1 bit, must include three additional 1 bits (so that a total of four 1 bits is sent), must include a total of 12 0 bits, and must have at least two 0 bits following each 1 bit?

32. How many different strings can be made from the letters in *MISSISSIPPI*, using all the letters?

33. How many different strings can be made from the letters in *ABRACADABRA*, using all the letters?

34. How many different strings can be made from the letters in *AARDVARK*, using all the letters, if all three *A*s must be consecutive?

35. How many different strings can be made from the letters in *ORONO*, using some or all of the letters?

36. How many strings with five or more characters can be formed from the letters in *SEERESS*?

37. How many strings with seven or more characters can be formed from the letters in *EVERGREEN*?

38. How many different bit strings can be formed using six 1s and eight 0s?

39. A student has three mangos, two papayas, and two kiwi fruits. If the student eats one piece of fruit each day, and only the type of fruit matters, in how many different ways can these fruits be consumed?

40. A professor packs her collection of 40 issues of a mathematics journal in four boxes with 10 issues per box. How many ways can she distribute the journals if

a) each box is numbered, so that they are distinguishable?

b) the boxes are identical, so that they cannot be distinguished?

41. How many ways are there to travel in xyz space from the origin $(0, 0, 0)$ to the point $(4, 3, 5)$ by taking steps one unit in the positive x direction, one unit in the positive y direction, or one unit in the positive z direction? (Moving in the negative x, y, or z direction is prohibited, so that no backtracking is allowed.)

42. How many ways are there to travel in $xyzw$ space from the origin $(0, 0, 0, 0)$ to the point $(4, 3, 5, 4)$ by taking steps one unit in the positive x, positive y, positive z, or positive w direction?

43. How many ways are there to deal hands of seven cards to each of five players from a standard deck of 52 cards?

44. In bridge, the 52 cards of a standard deck are dealt to four players. How many different ways are there to deal bridge hands to four players?

45. How many ways are there to deal hands of five cards to each of six players from a deck containing 48 different cards?

46. In how many ways can a dozen books be placed on four distinguishable shelves

a) if the books are indistinguishable copies of the same title?

b) if no two books are the same, and the positions of the books on the shelves matter? [*Hint:* Break this into 12 tasks, placing each book separately. Start with the sequence 1, 2, 3, 4 to represent the shelves. Represent the books by b_i, $i = 1, 2, \ldots, 12$. Place b_1 to the right of one of the terms in 1, 2, 3, 4. Then successively place $b_2, b_3, \ldots$, and b_{12}.]

47. How many ways can n books be placed on k distinguishable shelves

a) if the books are indistinguishable copies of the same title?

b) if no two books are the same, and the positions of the books on the shelves matter?

48. A shelf holds 12 books in a row. How many ways are there to choose five books so that no two adjacent books are chosen? [*Hint:* Represent the books that are chosen by bars and the books not chosen by stars. Count the number of sequences of five bars and seven stars so that no two bars are adjacent.]

***49.** Use the product rule to prove Theorem 4, by first placing objects in the first box, then placing objects in the second box, and so on.

***50.** Prove Theorem 4 by first setting up a one-to-one correspondence between permutations of n objects with n_i indistinguishable objects of type i, $i = 1, 2, 3, \ldots, k$, and the distributions of n objects in k boxes such that n_i objects are placed in box i, $i = 1, 2, 3, \ldots, k$ and then applying Theorem 3.

***51.** In this exercise we will prove Theorem 2 by setting up a one-to-one correspondence between the set of r-combinations with repetition allowed of $S = \{1, 2, 3, \ldots, n\}$ and the set of r-combinations of the set $T = \{1, 2, 3, \ldots, n + r - 1\}$.

a) Arrange the elements in an r-combination, with repetition allowed, of S into an increasing sequence $x_1 \le x_2 \le \cdots \le x_r$. Show that the sequence formed by adding $k - 1$ to the kth term is strictly increasing. Conclude that this sequence is made up of r distinct elements from T.

b) Show that the procedure described in (a) defines a one-to-one correspondence between the set of r-combinations, with repetition allowed, of S and the r-combinations of T. [*Hint:* Show the correspondence can be reversed by associating to the r-combination $\{x_1, x_2, \ldots, x_r\}$ of T, with $1 \le x_1 < x_2 < \cdots < x_r \le n + r - 1$, the r-combination with repetition allowed from S, formed by subtracting $k - 1$ from the kth element.]

c) Conclude that the number of r-combinations with repetition allowed from a set with n elements is $C(n + r - 1, r)$.

52. How many ways are there to distribute five distinguishable objects into three indistinguishable boxes?

53. How many ways are there to distribute six distinguishable objects into four indistinguishable boxes so that each of the boxes contains at least one object?

54. How many ways are there to put five temporary employees into four identical offices?

55. How many ways are there to put six temporary employees into four identical offices so that there is at least one temporary employee in each of these four offices?

56. How many ways are there to distribute five indistinguishable objects into three indistinguishable boxes?

57. How many ways are there to distribute six indistinguishable objects into four indistinguishable boxes so that each of the boxes contains at least one object?

58. How many ways are there to pack eight identical DVDs into five indistinguishable boxes so that each box contains at least one DVD?

59. How many ways are there to pack nine identical DVDs into three indistinguishable boxes so that each box contains at least two DVDs?

60. How many ways are there to distribute five balls into seven boxes if each box must have at most one ball in it if

a) both the balls and boxes are labeled?
b) the balls are labeled, but the boxes are unlabeled?
c) the balls are unlabeled, but the boxes are labeled?
d) both the balls and boxes are unlabeled?

61. How many ways are there to distribute five balls into three boxes if each box must have at least one ball in it if

a) both the balls and boxes are labeled?
b) the balls are labeled, but the boxes are unlabeled?
c) the balls are unlabeled, but the boxes are labeled?
d) both the balls and boxes are unlabeled?

62. Suppose that a basketball league has 32 teams, split into two conferences of 16 teams each. Each conference is split into three divisions. Suppose that the North Central Division has five teams. Each of the teams in the North Central Division plays four games against each of the other teams in this division, three games against each of the 11 remaining teams in the conference, and two games against each of the 16 teams in the other conference. In how many different orders can the games of one of the teams in the North Central Division be scheduled?

***63.** Suppose that a weapons inspector must inspect each of five different sites twice, visiting one site per day. The inspector is free to select the order in which to visit these sites, but cannot visit site X, the most suspicious site, on two consecutive days. In how many different orders can the inspector visit these sites?

64. How many different terms are there in the expansion of $(x_1 + x_2 + \cdots + x_m)^n$ after all terms with identical sets of exponents are added?

***65.** Prove the **Multinomial Theorem:** If n is a positive integer, then

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{n_1 + n_2 + \cdots + n_m = n} C(n; n_1, n_2, \ldots, n_m) x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m},$$

where

$$C(n; n_1, n_2, \ldots, n_m) = \frac{n!}{n_1!\, n_2! \cdots n_m!}$$

is a **multinomial coefficient**.

66. Find the expansion of $(x + y + z)^4$.

67. Find the coefficient of $x^3y^2z^5$ in $(x + y + z)^{10}$.

68. How many terms are there in the expansion of

$$(x + y + z)^{100}?$$

6.6 Generating Permutations and Combinations

6.6.1 Introduction

Methods for counting various types of permutations and combinations were described in the previous sections of this chapter, but sometimes permutations or combinations need to be generated, not just counted. Consider the following three problems. First, suppose that a salesperson must visit six different cities. In which order should these cities be visited to minimize total travel time? One way to determine the best order is to determine the travel time for each of the $6! = 720$ different orders in which the cities can be visited and choose the one with the smallest travel time. Second, suppose we are given a set of six positive integers and wish to find a subset of them that has 100 as their sum, if such a subset exists. One way to find these numbers is to generate all $2^6 = 64$ subsets and check the sum of their elements. Third, suppose a laboratory has 95 employees. A group of 12 of these employees with a particular set of 25 skills is needed for a project. (Each employee can have one or more of these skills.) One way to find such a set of employees is to generate all sets of 12 of these employees and check whether they have the desired skills. These examples show that it is often necessary to generate permutations and combinations to solve problems.

6.6.2 Generating Permutations

Links

Any set with n elements can be placed in one-to-one correspondence with the set $\{1, 2, 3, \ldots, n\}$. We can list the permutations of any set of n elements by generating the permutations of the n smallest positive integers and then replacing these integers with the corresponding elements. Many different algorithms have been developed to generate the $n!$ permutations of this set. We will describe one of these that is based on the **lexicographic** (or **dictionary**) **ordering** of the set of permutations of $\{1, 2, 3, \ldots, n\}$. In this ordering, the permutation $a_1a_2 \cdots a_n$ precedes the permutation of $b_1b_2 \cdots b_n$, if for some k, with $1 \leq k \leq n$, $a_1 = b_1, a_2 = b_2, \ldots, a_{k-1} = b_{k-1}$, and $a_k < b_k$. In other words, a permutation of the set of the n smallest positive integers precedes (in lexicographic order) a second permutation if the number in this permutation in the first position where the two permutations disagree is smaller than the number in that position in the second permutation.

EXAMPLE 1 The permutation 23415 of the set $\{1, 2, 3, 4, 5\}$ precedes the permutation 23514, because these permutations agree in the first two positions, but the number in the third position in the first permutation, 4, is smaller than the number in the third position in the second permutation, 5. Similarly, the permutation 41532 precedes 52143. ◀

An algorithm for generating the permutations of $\{1, 2, \ldots, n\}$ can be based on a procedure that constructs the next permutation in lexicographic order following a given permutation $a_1a_2 \cdots a_n$. We will show how this can be done. First, suppose that $a_{n-1} < a_n$. Interchange a_{n-1} and a_n to obtain a larger permutation. No other permutation is both larger than the original permutation and smaller than the permutation obtained by interchanging a_{n-1} and a_n. For instance, the next larger permutation after 234156 is 234165. On the other hand, if $a_{n-1} > a_n$, then a larger permutation cannot be obtained by interchanging these last two terms in the permutation. Look at the last three integers in the permutation. If $a_{n-2} < a_{n-1}$, then the last three integers in the permutation can be rearranged to obtain the next largest permutation. Put the smaller of the two integers a_{n-1} and a_n that is greater than a_{n-2} in position $n - 2$. Then, place the remaining integer and a_{n-2} into the last two positions in increasing order. For instance, the next larger permutation after 234165 is 234516.

On the other hand, if $a_{n-2} > a_{n-1}$ (and $a_{n-1} > a_n$), then a larger permutation cannot be obtained by permuting the last three terms in the permutation. Based on these observations, a general method can be described for producing the next larger permutation in increasing order following a given permutation $a_1a_2 \cdots a_n$. First, find the integers a_j and a_{j+1} with $a_j < a_{j+1}$ and

$$a_{j+1} > a_{j+2} > \cdots > a_n,$$

that is, the last pair of adjacent integers in the permutation where the first integer in the pair is smaller than the second. Then, the next larger permutation in lexicographic order is obtained by putting in the jth position the least integer among $a_{j+1}, a_{j+2}, \ldots$, and a_n that is greater than a_j and listing in increasing order the rest of the integers $a_j, a_{j+1}, \ldots, a_n$ in positions $j + 1$ to n. It is easy to see that there is no other permutation larger than the permutation $a_1a_2 \cdots a_n$ but smaller than the new permutation produced. (The verification of this fact is left as an exercise for the reader.)

EXAMPLE 2 What is the next permutation in lexicographic order after 362541?

Extra Examples

Solution: The last pair of integers a_j and a_{j+1} where $a_j < a_{j+1}$ is $a_3 = 2$ and $a_4 = 5$. The least integer to the right of 2 that is greater than 2 in the permutation is $a_5 = 4$. Hence, 4 is placed in the third position. Then the integers 2, 5, and 1 are placed in order in the last three positions, giving 125 as the last three positions of the permutation. Hence, the next permutation is 364125. ◀

To produce the $n!$ permutations of the integers $1, 2, 3, \ldots, n$, begin with the smallest permutation in lexicographic order, namely, $123 \cdots n$, and successively apply the procedure described for producing the next larger permutation of $n! - 1$ times. This yields all the permutations of the n smallest integers in lexicographic order.

EXAMPLE 3 Generate the permutations of the integers 1, 2, 3 in lexicographic order.

Solution: Begin with 123. The next permutation is obtained by interchanging 3 and 2 to obtain 132. Next, because $3 > 2$ and $1 < 3$, permute the three integers in 132. Put the smaller of 3 and 2 in the first position, and then put 1 and 3 in increasing order in positions 2 and 3 to obtain 213. This is followed by 231, obtained by interchanging 1 and 3, because $1 < 3$. The next larger permutation has 3 in the first position, followed by 1 and 2 in increasing order, namely, 312. Finally, interchange 1 and 2 to obtain the last permutation, 321. We have generated the permutations of 1, 2, 3 in lexicographic order. They are 123, 132, 213, 231, 312, and 321. ◀

Algorithm 1 displays the procedure for finding the next permutation in lexicographic order after a permutation that is not $n\ \ n-1\ \ n-2\ \ \ldots\ \ 2\ \ 1$, which is the largest permutation.

ALGORITHM 1 Generating the Next Permutation in Lexicographic Order.

procedure *next permutation*($a_1 a_2 \ldots a_n$: permutation of
  $\{1, 2, \ldots, n\}$ not equal to $n\ \ n-1\ \ \ldots\ \ 2\ \ 1$)
$j := n - 1$
while $a_j > a_{j+1}$
  $j := j - 1$
{j is the largest subscript with $a_j < a_{j+1}$}
$k := n$
while $a_j > a_k$
  $k := k - 1$
{a_k is the smallest integer greater than a_j to the right of a_j}
interchange a_j and a_k
$r := n$
$s := j + 1$
while $r > s$
  interchange a_r and a_s
  $r := r - 1$
  $s := s + 1$
{this puts the tail end of the permutation after the jth position in increasing order}
{$a_1 a_2 \ldots a_n$ is now the next permutation}

6.6.3 Generating Combinations

Links

How can we generate all the combinations of the elements of a finite set? Because a combination is just a subset, we can use the correspondence between subsets of $\{a_1, a_2, \ldots, a_n\}$ and bit strings of length n.

Recall that the bit string corresponding to a subset has a 1 in position k if a_k is in the subset, and has a 0 in this position if a_k is not in the subset. If all the bit strings of length n can be listed, then by the correspondence between subsets and bit strings, a list of all the subsets is obtained.

Recall that a bit string of length n is also the binary expansion of an integer between 0 and $2^n - 1$. The 2^n bit strings can be listed in order of their increasing size as integers in their binary expansions. To produce all binary expansions of length n, start with the bit string $000 \ldots 00$, with n zeros. Then, successively find the next expansion until the bit string $111 \ldots 11$ is obtained. At each stage the next binary expansion is found by locating the first position from the right that is not a 1, then changing all the 1s to the right of this position to 0s and making this first 0 (from the right) a 1.

EXAMPLE 4 Find the next bit string after 10 0010 0111.

Solution: The first bit from the right that is not a 1 is the fourth bit from the right. Change this bit to a 1 and change all the following bits to 0s. This produces the next larger bit string, 10 0010 1000. ◀

The procedure for producing the next larger bit string after $b_{n-1}b_{n-2}\ldots b_1b_0$ is given as Algorithm 2.

ALGORITHM 2 Generating the Next Larger Bit String.

```
procedure next bit string(b_{n−1} b_{n−2}...b_1b_0: bit string not equal to 11...11)
i := 0
while b_i = 1
   b_i := 0
   i := i + 1
b_i := 1
{b_{n−1} b_{n−2}...b_1b_0 is now the next bit string}
```

Next, an algorithm for generating the r-combinations of the set $\{1, 2, 3, \ldots, n\}$ will be given. An r-combination can be represented by a sequence containing the elements in the subset in increasing order. The r-combinations can be listed using lexicographic order on these sequences. In this lexicographic ordering, the first r-combination is $\{1, 2, \ldots, r-1, r\}$ and the last r-combination is $\{n-r+1, n-r+2, \ldots, n-1, n\}$. The next r-combination after $a_1a_2 \cdots a_r$ can be obtained in the following way: First, locate the last element a_i in the sequence such that $a_i \neq n-r+i$. Then, replace a_i with $a_i + 1$ and a_j with $a_i + j - i + 1$, for $j = i+1, i+2, \ldots, r$. It is left for the reader to show that this produces the next larger r-combination in lexicographic order. This procedure is illustrated with Example 5.

EXAMPLE 5 Find the next larger 4-combination of the set $\{1, 2, 3, 4, 5, 6\}$ after $\{1, 2, 5, 6\}$.

Solution: The last term among the terms a_i with $a_1 = 1$, $a_2 = 2$, $a_3 = 5$, and $a_4 = 6$ such that $a_i \neq 6 - 4 + i$ is $a_2 = 2$. To obtain the next larger 4-combination, increment a_2 by 1 to obtain $a_2 = 3$. Then set $a_3 = 3 + 1 = 4$ and $a_4 = 3 + 2 = 5$. Hence the next larger 4-combination is $\{1, 3, 4, 5\}$. ◀

Algorithm 3 displays pseudocode for this procedure.

ALGORITHM 3 Generating the Next r-Combination in Lexicographic Order.

```
procedure next r-combination({a_1, a_2, ..., a_r}: proper subset of
        {1, 2, ..., n} not equal to {n − r + 1, ..., n} with
        a_1 < a_2 < ··· < a_r)
i := r
while a_i = n − r + i
   i := i − 1
a_i := a_i + 1
for j := i + 1 to r
   a_j := a_i + j − i
{{a_1, a_2, ..., a_r} is now the next combination}
```

Exercises

1. Place these permutations of $\{1, 2, 3, 4, 5\}$ in lexicographic order: 43521, 15432, 45321, 23451, 23514, 14532, 21345, 45213, 31452, 31542.
2. Place these permutations of $\{1,2,3,4,5,6\}$ in lexicographic order: 234561, 231456, 165432, 156423, 543216, 541236, 231465, 314562, 432561, 654321, 654312, 435612.
3. The name of a file in a computer directory consists of three uppercase letters followed by a digit, where each letter is either A, B, or C, and each digit is either 1 or 2. List the name of these files in lexicographic order, where we order letters using the usual alphabetic order of letters.
4. Suppose that the name of a file in a computer directory consists of three digits followed by two lowercase letters and each digit is 0, 1, or 2, and each letter is either a or b. List the name of these files in lexicographic order, where we order letters using the usual alphabetic order of letters.
5. Find the next larger permutation in lexicographic order after each of these permutations.
   a) 1432 **b)** 54123 **c)** 12453
   d) 45231 **e)** 6714235 **f)** 31528764
6. Find the next larger permutation in lexicographic order after each of these permutations.
   a) 1342 **b)** 45321 **c)** 13245
   d) 612345 **e)** 1623547 **f)** 23587416
7. Use Algorithm 1 to generate the 24 permutations of the first four positive integers in lexicographic order.
8. Use Algorithm 2 to list all the subsets of the set $\{1, 2, 3, 4\}$.
9. Use Algorithm 3 to list all the 3-combinations of $\{1, 2, 3, 4, 5\}$.
10. Show that Algorithm 1 produces the next larger permutation in lexicographic order.
11. Show that Algorithm 3 produces the next larger r-combination in lexicographic order after a given r-combination.
12. Develop an algorithm for generating the r-permutations of a set of n elements.
13. List all 3-permutations of $\{1, 2, 3, 4, 5\}$.

The remaining exercises in this section develop another algorithm for generating the permutations of $\{1, 2, 3, \ldots, n\}$. This algorithm is based on Cantor expansions of integers. Every nonnegative integer less than $n!$ has a unique Cantor expansion

$$a_1 1! + a_2 2! + \cdots + a_{n-1}(n-1)!$$

where a_i is a nonnegative integer not exceeding i, for $i = 1, 2, \ldots, n-1$. The integers $a_1, a_2, \ldots, a_{n-1}$ are called the **Cantor digits** of this integer.

Given a permutation of $\{1, 2, \ldots, n\}$, let a_{k-1}, $k = 2, 3, \ldots, n$, be the number of integers less than k that follow k in the permutation. For instance, in the permutation 43215, a_1 is the number of integers less than 2 that follow 2, so $a_1 = 1$. Similarly, for this example $a_2 = 2$, $a_3 = 3$, and $a_4 = 0$. Consider the function from the set of permutations of $\{1, 2, 3, \ldots, n\}$ to the set of nonnegative integers less than $n!$ that sends a permutation to the integer that has $a_1, a_2, \ldots, a_{n-1}$, defined in this way, as its Cantor digits.

14. Find the Cantor digits $a_1, a_2, \ldots, a_{n-1}$ that correspond to these permutations.
    a) 246531 **b)** 12345 **c)** 654321

*15. Show that the correspondence described in the preamble is a bijection between the set of permutations of $\{1, 2, 3, \ldots, n\}$ and the nonnegative integers less than $n!$.

16. Find the permutations of $\{1, 2, 3, 4, 5\}$ that correspond to these integers with respect to the correspondence between Cantor expansions and permutations as described in the preamble to Exercise 14.
    a) 3 **b)** 89 **c)** 111
17. Develop an algorithm for producing all permutations of a set of n elements based on the correspondence described in the preamble to Exercise 14.

Key Terms and Results

TERMS

combinatorics: the study of arrangements of objects

enumeration: the counting of arrangements of objects

tree diagram: a diagram made up of a root, branches leaving the root, and other branches leaving some of the endpoints of branches

permutation: an ordered arrangement of the elements of a set

***r*-permutation:** an ordered arrangement of r elements of a set

$P(n,r)$: the number of r-permutations of a set with n elements

***r*-combination:** an unordered selection of r elements of a set

$C(n,r)$: the number of r-combinations of a set with n elements

binomial coefficient $\binom{n}{r}$: also the number of r-combinations of a set with n elements

combinatorial proof: a proof that uses counting arguments rather than algebraic manipulation to prove a result

Pascal's triangle: a representation of the binomial coefficients where the ith row of the triangle contains $\binom{i}{j}$ for $j = 0, 1, 2, \ldots, i$

$S(n,j)$: the Stirling number of the second kind denoting the number of ways to distribute n distinguishable objects into j indistinguishable boxes so that no box is empty

RESULTS

product rule for counting: The number of ways to do a procedure that consists of two tasks is the product of the number of ways to do the first task and the number of ways to do the second task after the first task has been done.

product rule for sets: The number of elements in the Cartesian product of finite sets is the product of the number of elements in each set.

sum rule for counting: The number of ways to do a task in one of two ways is the sum of the number of ways to do these tasks if they cannot be done simultaneously.

sum rule for sets: The number of elements in the union of pairwise disjoint finite sets is the sum of the numbers of elements in these sets.

subtraction rule for counting or **inclusion–exclusion for sets:** If a task can be done in either n_1 ways or n_2 ways, then the number of ways to do the task is $n_1 + n_2$ minus the number of ways to do the task that are common to the two different ways.

subtraction rule or **inclusion–exclusion for sets:** The number of elements in the union of two sets is the sum of the number of elements in these sets minus the number of elements in their intersection.

division rule for counting: There are n/d ways to do a task if it can be done using a procedure that can be carried out in n ways, and for every way w, exactly d of the n ways correspond to way w.

division rule for sets: Suppose that a finite set A is the union of n disjoint subsets each with d elements. Then $n = |A|/d$.

the pigeonhole principle: When more than k objects are placed in k boxes, there must be a box containing more than one object.

the generalized pigeonhole principle: When N objects are placed in k boxes, there must be a box containing at least $\lceil N/k \rceil$ objects.

$$P(n, r) = \frac{n!}{(n-r)!}$$

$$C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Pascal's identity: $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$

the binomial theorem: $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$

There are n^r r-permutations of a set with n elements when repetition is allowed.

There are $C(n + r - 1, r)$ r-combinations of a set with n elements when repetition is allowed.

There are $n!/(n_1!\, n_2! \cdots n_k!)$ permutations of n objects of k types where there are n_i indistinguishable objects of type i for $i = 1, 2, 3, \ldots, k$.

the algorithm for generating the permutations of the set $\{1, 2, \ldots, n\}$

Review Questions

1. Explain how the sum and product rules can be used to find the number of bit strings with a length not exceeding 10.

2. Explain how to find the number of bit strings of length not exceeding 10 that have at least one 0 bit.

3. **a)** How can the product rule be used to find the number of functions from a set with m elements to a set with n elements?
   b) How many functions are there from a set with five elements to a set with 10 elements?
   c) How can the product rule be used to find the number of one-to-one functions from a set with m elements to a set with n elements?
   d) How many one-to-one functions are there from a set with five elements to a set with 10 elements?
   e) How many onto functions are there from a set with five elements to a set with 10 elements?

4. How can you find the number of possible outcomes of a playoff between two teams where the first team that wins four games wins the playoff?

5. How can you find the number of bit strings of length ten that either begin with 101 or end with 010?

6. **a)** State the pigeonhole principle.
   b) Explain how the pigeonhole principle can be used to show that among any 11 integers, at least two must have the same last digit.

7. **a)** State the generalized pigeonhole principle.
   b) Explain how the generalized pigeonhole principle can be used to show that among any 91 integers, there are at least ten that end with the same digit.

8. **a)** What is the difference between an r-combination and an r-permutation of a set with n elements?
   b) Derive an equation that relates the number of r-combinations and the number of r-permutations of a set with n elements.
   c) How many ways are there to select six students from a class of 25 to serve on a committee?
   d) How many ways are there to select six students from a class of 25 to hold six different executive positions on a committee?

9. **a)** What is Pascal's triangle?
   b) How can a row of Pascal's triangle be produced from the one above it?

10. What is meant by a combinatorial proof of an identity? How is such a proof different from an algebraic one?

11. Explain how to prove Pascal's identity using a combinatorial argument.

12. a) State the binomial theorem.
 b) Explain how to prove the binomial theorem using a combinatorial argument.
 c) Find the coefficient of $x^{100}y^{101}$ in the expansion of $(2x + 5y)^{201}$.
13. a) Explain how to find a formula for the number of ways to select r objects from n objects when repetition is allowed and order does not matter.
 b) How many ways are there to select a dozen objects from among objects of five different types if objects of the same type are indistinguishable?
 c) How many ways are there to select a dozen objects from these five different types if there must be at least three objects of the first type?
 d) How many ways are there to select a dozen objects from these five different types if there cannot be more than four objects of the first type?
 e) How many ways are there to select a dozen objects from these five different types if there must be at least two objects of the first type, but no more than three objects of the second type?
14. a) Let n and r be positive integers. Explain why the number of solutions of the equation $x_1 + x_2 + \cdots + x_n = r$, where x_i is a nonnegative integer for $i = 1, 2, 3, \ldots, n$, equals the number of r-combinations of a set with n elements.
 b) How many solutions in nonnegative integers are there to the equation $x_1 + x_2 + x_3 + x_4 = 17$?
 c) How many solutions in positive integers are there to the equation in part (b)?
15. a) Derive a formula for the number of permutations of n objects of k different types, where there are n_1 indistinguishable objects of type one, n_2 indistinguishable objects of type two, ..., and n_k indistinguishable objects of type k.
 b) How many ways are there to order the letters of the word *INDISCREETNESS*?
16. Describe an algorithm for generating all the permutations of the set of the n smallest positive integers.
17. a) How many ways are there to deal hands of five cards to six players from a standard 52-card deck?
 b) How many ways are there to distribute n distinguishable objects into k distinguishable boxes so that n_i objects are placed in box i?
18. Describe an algorithm for generating all the combinations of the set of the n smallest positive integers.

Supplementary Exercises

1. How many ways are there to choose 6 items from 10 distinct items when
 a) the items in the choices are ordered and repetition is not allowed?
 b) the items in the choices are ordered and repetition is allowed?
 c) the items in the choices are unordered and repetition is not allowed?
 d) the items in the choices are unordered and repetition is allowed?
2. How many ways are there to choose 10 items from 6 distinct items when
 a) the items in the choices are ordered and repetition is not allowed?
 b) the items in the choices are ordered and repetition is allowed?
 c) the items in the choices are unordered and repetition is not allowed?
 d) the items in the choices are unordered and repetition is allowed?
3. A test contains 100 true/false questions. How many different ways can a student answer the questions on the test, if answers may be left blank?
4. How many strings of length 10 either start with 000 or end with 1111?
5. How many bit strings of length 10 over the alphabet $\{a, b, c\}$ have either exactly three as or exactly four bs?
6. The internal telephone numbers in the phone system on a campus consist of five digits, with the first digit not equal to zero. How many different numbers can be assigned in this system?
7. An ice cream parlor has 28 different flavors, 8 different kinds of sauce, and 12 toppings.
 a) In how many different ways can a dish of three scoops of ice cream be made where each flavor can be used more than once and the order of the scoops does not matter?
 b) How many different kinds of small sundaes are there if a small sundae contains one scoop of ice cream, a sauce, and a topping?
 c) How many different kinds of large sundaes are there if a large sundae contains three scoops of ice cream, where each flavor can be used more than once and the order of the scoops does not matter; two kinds of sauce, where each sauce can be used only once and the order of the sauces does not matter; and three toppings, where each topping can be used only once and the order of the toppings does not matter?
8. How many positive integers less than 1000
 a) have exactly three decimal digits?
 b) have an odd number of decimal digits?
 c) have at least one decimal digit equal to 9?
 d) have no odd decimal digits?
 e) have two consecutive decimal digits equal to 5?
 f) are palindromes (that is, read the same forward and backward)?

9. When the numbers from 1 to 1000 are written out in decimal notation, how many of each of these digits are used?

a) 0 **b)** 1 **c)** 2 **d)** 9

10. There are 12 signs of the zodiac. How many people are needed to guarantee that at least six of these people have the same sign?

11. A fortune cookie company makes 213 different fortunes. A student eats at a restaurant that uses fortunes from this company and gives each customer one fortune cookie at the end of a meal. What is the largest possible number of times that the student can eat at the restaurant without getting the same fortune four times?

12. How many people are needed to guarantee that at least two were born on the same day of the week and in the same month (perhaps in different years)?

13. Show that given any set of 10 positive integers not exceeding 50 there exist at least two different five-element subsets of this set that have the same sum.

14. A package of baseball cards contains 20 cards. How many packages must be purchased to ensure that two cards in these packages are identical if there are a total of 550 different cards?

15. **a)** How many cards must be chosen from a standard deck of 52 cards to guarantee that at least two of the four aces are chosen?

b) How many cards must be chosen from a standard deck of 52 cards to guarantee that at least two of the four aces and at least two of the 13 kinds are chosen?

c) How many cards must be chosen from a standard deck of 52 cards to guarantee that there are at least two cards of the same kind?

d) How many cards must be chosen from a standard deck of 52 cards to guarantee that there are at least two cards of each of two different kinds?

***16.** Show that in any set of $n + 1$ positive integers not exceeding $2n$ there must be two that are relatively prime.

***17.** Show that in a sequence of m integers there exists one or more consecutive terms with a sum divisible by m.

18. Show that if five points are picked in the interior of a square with a side length of 2, then at least two of these points are no farther than $\sqrt{2}$ apart.

19. Show that the decimal expansion of a rational number must repeat itself from some point onward.

20. Once a computer worm infects a personal computer via an infected e-mail message, it sends a copy of itself to 100 e-mail addresses it finds in the electronic message mailbox on this personal computer. What is the maximum number of different computers this one computer can infect in the time it takes for the infected message to be forwarded five times?

21. How many ways are there to choose a dozen donuts from 20 varieties

a) if there are no two donuts of the same variety?

b) if all donuts are of the same variety?

c) if there are no restrictions?

d) if there are at least two varieties among the dozen donuts chosen?

e) if there must be at least six blueberry-filled donuts?

f) if there can be no more than six blueberry-filled donuts?

22. Find n if

a) $P(n, 2) = 110$. **b)** $P(n, n) = 5040$.

c) $P(n, 4) = 12P(n, 2)$.

23. Find n if

a) $C(n, 2) = 45$. **b)** $C(n, 3) = P(n, 2)$.

c) $C(n, 5) = C(n, 2)$.

24. Show that if n and r are nonnegative integers and $n \geq r$, then

$$P(n + 1, r) = P(n, r)(n + 1)/(n + 1 - r).$$

***25.** Suppose that S is a set with n elements. How many ordered pairs (A, B) are there such that A and B are subsets of S with $A \subseteq B$? [*Hint:* Show that each element of S belongs to A, $B - A$, or $S - B$.]

26. Give a combinatorial proof of Corollary 2 of Section 6.4 by setting up a correspondence between the subsets of a set with an even number of elements and the subsets of this set with an odd number of elements. [*Hint:* Take an element a in the set. Set up the correspondence by putting a in the subset if it is not already in it and taking it out if it is in the subset.]

27. Let n and r be integers with $1 \leq r < n$. Show that

$$C(n, r - 1) = C(n + 2, r + 1) - 2C(n + 1, r + 1) + C(n, r + 1).$$

28. Prove using mathematical induction that $\sum_{j=2}^{n} C(j, 2) = C(n + 1, 3)$ whenever n is an integer greater than 1.

29. Show that if n is an integer then

$$\sum_{k=0}^{n} 3^k \binom{n}{k} = 4^n.$$

30. Show that $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = \binom{n}{2}$ if n is an integer with $n \geq 2$.

31. Show that $\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} 1 = \binom{n}{3}$ if n is an integer with $n \geq 3$.

32. In this exercise we will derive a formula for the sum of the squares of the n smallest positive integers. We will count the number of triples (i, j, k) where i, j, and k are integers such that $0 \leq i < k$, $0 \leq j < k$, and $1 \leq k \leq n$ in two ways.

a) Show that there are k^2 such triples with a fixed k. Deduce that there are $\sum_{k=1}^{n} k^2$ such triples.

b) Show that the number of such triples with $0 \leq i < j < k$ and the number of such triples with $0 \leq j < i < k$ both equal $C(n + 1, 3)$.

c) Show that the number of such triples with $0 \leq i = j < k$ equals $C(n + 1, 2)$.

d) Combining part (a) with parts (b) and (c), conclude that

$$\sum_{k=1}^{n} k^2 = 2C(n+1,3) + C(n+1,2)$$
$$= n(n+1)(2n+1)/6.$$

*** 33.** How many bit strings of length n, where $n \geq 4$, contain exactly two occurrences of 01?

34. Let S be a set. We say that a collection of subsets $A_1, A_2, \ldots, A_n$ each containing d elements, where $d \geq 2$, is 2-*colorable* if it is possible to assign to each element of S one of two different colors so that in every subset A_i there are elements that have been assigned each color. Let $m(d)$ be the largest integer such that every collection of fewer than $m(d)$ sets each containing d elements is 2-colorable.

a) Show that the collection of all subsets with d elements of a set S with $2d - 1$ elements is not 2-colorable.

b) Show that $m(2) = 3$.

****c)** Show that $m(3) = 7$. [*Hint:* Show that the collection $\{1, 3, 5\}$, $\{1, 2, 6\}$, $\{1, 4, 7\}$, $\{2, 3, 4\}$, $\{2, 5, 7\}$, $\{3, 6, 7\}$, $\{4, 5, 6\}$ is not 2-colorable. Then show that all collections of six sets with three elements each are 2-colorable.]

35. A professor writes 20 multiple-choice questions, each with the possible answer a, b, c, or d, for a discrete mathematics test. If the number of questions with a, b, c, and d as their answer is 8, 3, 4, and 5, respectively, how many different answer keys are possible, if the questions can be placed in any order?

36. How many different arrangements are there of eight people seated at a round table, where two arrangements are considered the same if one can be obtained from the other by a rotation?

37. How many ways are there to assign 24 students to five faculty advisors?

38. How many ways are there to choose a dozen apples from a bushel containing 20 indistinguishable Delicious apples, 20 indistinguishable Macintosh apples, and 20 indistinguishable Granny Smith apples, if at least three of each kind must be chosen?

39. How many solutions are there to the equation $x_1 + x_2 + x_3 = 17$, where x_1, x_2, and x_3 are nonnegative integers with

a) $x_1 > 1$, $x_2 > 2$, and $x_3 > 3$?

b) $x_1 < 6$ and $x_3 > 5$?

c) $x_1 < 4$, $x_2 < 3$, and $x_3 > 5$?

40. a) How many different strings can be made from the word *PEPPERCORN* when all the letters are used?

b) How many of these strings start and end with the letter *P*?

c) In how many of these strings are the three letter *P*s consecutive?

41. How many subsets of a set with ten elements

a) have fewer than five elements?

b) have more than seven elements?

c) have an odd number of elements?

42. A witness to a hit-and-run accident tells the police that the license plate of the car in the accident, which contains three letters followed by three digits, starts with the letters *AS* and contains both the digits 1 and 2. How many different license plates can fit this description?

43. How many ways are there to put n identical objects into m distinct containers so that no container is empty?

44. How many ways are there to seat six boys and eight girls in a row of chairs so that no two boys are seated next to each other?

45. How many ways are there to distribute six objects to five boxes if

a) both the objects and boxes are labeled?

b) the objects are labeled, but the boxes are unlabeled?

c) the objects are unlabeled, but the boxes are labeled?

d) both the objects and the boxes are unlabeled?

46. How many ways are there to distribute five objects into six boxes if

a) both the objects and boxes are labeled?

b) the objects are labeled, but the boxes are unlabeled?

c) the objects are unlabeled, but the boxes are labeled?

d) both the objects and the boxes are unlabeled?

The **signless Stirling number of the first kind** $c(n, k)$, where k and n are integers with $1 \leq k \leq n$, equals the number of ways to arrange n people around k circular tables with at least one person seated at each table, where two seatings of m people around a circular table are considered the same if everyone has the same left neighbor and the same right neighbor.

47. Find these signless Stirling numbers of the first kind.

a) $c(3, 2)$ **b)** $c(4, 2)$

c) $c(4, 3)$ **d)** $c(5, 4)$

48. Show that if n is a positive integer, then $\sum_{j=1}^{n} c(n, j) = n!$.

49. Show that if n is a positive integer with $n \geq 3$, then $c(n, n-2) = (3n-1)C(n, 3)/4$.

*** 50.** Show that if n and k are integers with $1 \leq k < n$, then $c(n+1, k) = c(n, k-1) + nc(n, k)$.

51. Give a combinatorial proof that 2^n divides $n!$ whenever n is an even positive integer. [*Hint:* Use Theorem 3 in Section 6.5 to count the number of permutations of $2n$ objects where there are two indistinguishable objects of n different types.]

52. How many 11-element RNA sequences consist of 4 As, 3 Cs, 2 Us, and 2 Gs, and end with CAA?

Exercises 53 and 54 are based on a discussion in [RoTe09]. A method used in the 1960s for sequencing RNA chains used enzymes to break chains after certain links. Some enzymes break RNA chains after each G link, while others break them after each C or U link. Using these enzymes it is sometimes possible to correctly sequence all the bases in an RNA chain.

*** 53.** Suppose that when an enzyme that breaks RNA chains after each G link is applied to a 12-link chain, the fragments obtained are G, CCG, AAAG, and UCCG, and when an enzyme that breaks RNA chains after each C or U link is applied, the fragments obtained are C, C, C, C, GGU, and GAAAG. Can you determine the entire 12-link RNA chain from these two sets of fragments? If so, what is this RNA chain?

*54. Suppose that when an enzyme that breaks RNA chains after each G link is applied to a 12-link chain, the fragments obtained are AC, UG, and ACG and when an enzyme that breaks RNA chains after each C or U link is applied, the fragments obtained are U, GAC, and GAC. Can you determine the entire RNA chain from these two sets of fragments? If so, what is this RNA chain?

55. Devise an algorithm for generating all the r-permutations of a finite set when repetition is allowed.

56. Devise an algorithm for generating all the r-combinations of a finite set when repetition is allowed.

*57. Show that if m and n are integers with $m \geq 3$ and $n \geq 3$, then $R(m, n) \leq R(m, n-1) + R(m-1, n)$.

*58. Show that $R(3, 4) \geq 7$ by showing that in a group of six people, where any two people are friends or enemies, there are not necessarily three mutual friends or four mutual enemies.

Computer Projects

Write programs with these input and output.

1. Given a positive integer n and a nonnegative integer not exceeding n, find the number of r-permutations and r-combinations of a set with n elements.
2. Given positive integers n and r, find the number of r-permutations when repetition is allowed and r-combinations when repetition is allowed of a set with n elements.
3. Given a sequence of positive integers, find the longest increasing and the longest decreasing subsequence of the sequence.

*4. Given an equation $x_1 + x_2 + \cdots + x_n = C$, where C is a constant, and $x_1, x_2, \ldots, x_n$ are nonnegative integers, list all the solutions.

5. Given a positive integer n, list all the permutations of the set $\{1, 2, 3, \ldots, n\}$ in lexicographic order.
6. Given a positive integer n and a nonnegative integer r not exceeding n, list all the r-combinations of the set $\{1, 2, 3, \ldots, n\}$ in lexicographic order.
7. Given a positive integer n and a nonnegative integer r not exceeding n, list all the r-permutations of the set $\{1, 2, 3, \ldots, n\}$ in lexicographic order.
8. Given a positive integer n, list all the combinations of the set $\{1, 2, 3, \ldots, n\}$.
9. Given positive integers n and r, list all the r-permutations, with repetition allowed, of the set $\{1, 2, 3, \ldots, n\}$.
10. Given positive integers n and r, list all the r-combinations, with repetition allowed, of the set $\{1, 2, 3, \ldots, n\}$.

Computations and Explorations

Use a computational program or programs you have written to do these exercises.

1. Find the number of possible outcomes in a two-team playoff when the winner is the first team to win 5 out of 9, 6 out of 11, 7 out of 13, and 8 out of 15.
2. Which binomial coefficients are odd? Can you formulate a conjecture based on numerical evidence?
3. Verify that $C(2n, n)$ is divisible by the square of a prime, when $n \neq 1, 2,$ or 4, for as many positive integers n as you can. [That $C(2n, n)$ is divisible by the square of a prime with $n \neq 1, 2,$ or 4 was proved in 1996 by Andrew Granville and Olivier Ramaré, settling a conjecture made in 1980 by Paul Erdős and Ron Graham.]
4. Find as many odd integers n less than 200 as you can for which $C(n, \lfloor n/2 \rfloor)$ is not divisible by the square of a prime. Formulate a conjecture based on your evidence.

*5. For each integer less than 100 determine whether $C(2n, n)$ is divisible by 3. Can you formulate a conjecture that tells us for which integers n the binomial coefficient $C(2n, n)$ is divisible by 3 based on the digits in the base three expansion of n?

6. Generate all the permutations of a set with eight elements.
7. Generate all the 6-permutations of a set with nine elements.
8. Generate all combinations of a set with eight elements.
9. Generate all 5-combinations with repetition allowed of a set with seven elements.

Writing Projects

Respond to these with essays using outside sources.

1. Describe some of the earliest uses of the pigeonhole principle by Dirichlet and other mathematicians.
2. Discuss ways in which the current telephone numbering plan can be extended to accommodate the rapid demand for more telephone numbers. (See if you can find some of the proposals coming from the telecommunications industry.) For each new numbering plan you discuss, show how to find the number of different telephone numbers it supports.
3. Discuss the importance of combinatorial reasoning in gene sequencing and related problems involving genomes.
4. Many combinatorial identities are described in this book. Find some sources of such identities and describe important combinatorial identities besides those already introduced in this book. Give some representative proofs, including combinatorial ones, of some of these identities.
5. Describe the different models used to model the distribution of particles in statistical mechanics, including Maxwell–Boltzmann, Bose–Einstein, and Fermi–Dirac statistics. In each case, describe the counting techniques used in the model.
6. Define the Stirling numbers of the first kind and describe some of their properties and the identities they satisfy.
7. Describe some of the properties and the identities that Stirling numbers of the second kind satisfy, including the connection between Stirling numbers of the first and second kinds.
8. Describe the latest discoveries of values and bounds for Ramsey numbers.
9. Describe additional ways to generate all the permutations of a set with n elements besides those found in Section 6.6. Compare these algorithms and the algorithms described in the text and exercises of Section 6.6 in terms of their computational complexity.
10. Describe at least one way to generate all the partitions of a positive integer n. (See Exercise 49 in Section 5.3.)

CHAPTER

8 Advanced Counting Techniques

8.1 Applications of Recurrence Relations

8.2 Solving Linear Recurrence Relations

8.3 Divide-and-Conquer Algorithms and Recurrence Relations

8.4 Generating Functions

8.5 Inclusion–Exclusion

8.6 Applications of Inclusion–Exclusion

Many counting problems cannot be solved easily using the methods discussed in Chapter 6. One such problem is: How many bit strings of length n do not contain two consecutive zeros? To solve this problem, let a_n be the number of such strings of length n. An argument can be given that shows that the sequence $\{a_n\}$ satisfies the recurrence relation $a_{n+1} = a_n + a_{n-1}$ and the initial conditions $a_1 = 2$ and $a_2 = 3$. This recurrence relation and the initial conditions determine the sequence $\{a_n\}$. Moreover, an explicit formula can be found for a_n from the equation relating the terms of the sequence. As we will see, a similar technique can be used to solve many different types of counting problems.

We will discuss two ways that recurrence relations play important roles in the study of algorithms. First, we will introduce an important algorithmic paradigm known as dynamic programming. Algorithms that follow this paradigm break down a problem into overlapping subproblems. The solution to the problem is then found from the solutions to the subproblems through the use of a recurrence relation. Second, we will study another important algorithmic paradigm, divide-and-conquer. Algorithms that follow this paradigm can be used to solve a problem by recursively breaking it into a fixed number of nonoverlapping subproblems until these problems can be solved directly. The complexity of such algorithms can be analyzed using a special type of recurrence relation. In this chapter we will discuss a variety of divide-and-conquer algorithms and analyze their complexity using recurrence relations.

We will also see that many counting problems can be solved using formal power series, called generating functions, where the coefficients of powers of x represent terms of the sequence we are interested in. Besides solving counting problems, we will also be able to use generating functions to solve recurrence relations and to prove combinatorial identities.

Many other kinds of counting problems cannot be solved using the techniques discussed in Chapter 6, such as: How many ways are there to assign seven jobs to three employees so that each employee is assigned at least one job? How many primes are there less than 1000? Both of these problems can be solved by counting the number of elements in the union of sets. We will develop a technique, called the principle of inclusion–exclusion, that counts the number of elements in a union of sets, and we will show how this principle can be used to solve counting problems.

The techniques studied in this chapter, together with the basic techniques of Chapter 6, can be used to solve many counting problems.

8.1 Applications of Recurrence Relations

8.1.1 Introduction

Recall from Chapter 2 that a recursive definition of a sequence specifies one or more initial terms and a rule for determining subsequent terms from those that precede them. Also, recall that a rule of the latter sort (whether or not it is part of a recursive definition) is called a **recurrence relation** and that a sequence is called a *solution* of a recurrence relation if its terms satisfy the recurrence relation.

In this section we will show that such relations can be used to study and to solve counting problems. For example, suppose that the number of bacteria in a colony doubles every hour. If a colony begins with five bacteria, how many will be present in n hours? To solve this problem, let a_n be the number of bacteria at the end of n hours. Because the number of bacteria doubles every hour, the relationship $a_n = 2a_{n-1}$ holds whenever n is a positive integer. This recurrence relation, together with the initial condition $a_0 = 5$, uniquely determines a_n for all nonnegative integers n. We can find a formula for a_n using the iterative approach followed in Chapter 2, namely that $a_n = 5 \cdot 2^n$ for all nonnegative integers n.

Some of the counting problems that cannot be solved using the techniques discussed in Chapter 6 can be solved by finding recurrence relations involving the terms of a sequence, as was done in the problem involving bacteria. In this section we will study a variety of counting problems that can be modeled using recurrence relations. In Chapter 2 we developed methods for solving certain recurrence relation. In Section 8.2 we will study methods for finding explicit formulae for the terms of sequences that satisfy certain types of recurrence relations.

We conclude this section by introducing the algorithmic paradigm of dynamic programming. After explaining how this paradigm works, we will illustrate its use with an example.

8.1.2 Modeling With Recurrence Relations

Assessment

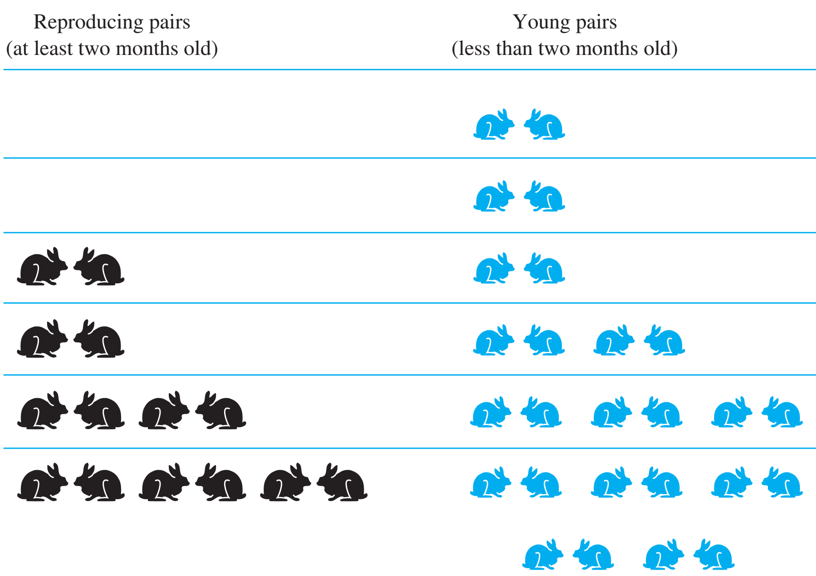
We can use recurrence relations to model a wide variety of problems, such as finding compound interest (see Example 11 in Section 2.4), counting rabbits on an island, determining the number of moves in the Tower of Hanoi puzzle, and counting bit strings with certain properties.

Extra Examples

Example 1 shows how the population of rabbits on an island can be modeled using a recurrence relation.

EXAMPLE 1 **Rabbits and the Fibonacci Numbers** Consider this problem, which was originally posed by Leonardo Pisano, also known as Fibonacci, in the thirteenth century in his book *Liber abaci*. A young pair of rabbits (one of each sex) is placed on an island. A pair of rabbits does not breed until they are 2 months old. After they are 2 months old, each pair of rabbits produces another pair each month, as shown in Figure 1. Find a recurrence relation for the number of pairs of rabbits on the island after n months, assuming that no rabbits ever die.

Links



Month	Reproducing pairs	Young pairs	Total pairs
1	0	1	1
2	0	1	1
3	1	1	2
4	1	2	3
5	2	3	5
6	3	5	8

FIGURE 1 **Rabbits on an island.**

Solution: Denote by f_n the number of pairs of rabbits after n months. We will show that f_n, $n = 1, 2, 3, \ldots$, are the terms of the Fibonacci sequence.

The rabbit population can be modeled using a recurrence relation. At the end of the first month, the number of pairs of rabbits on the island is $f_1 = 1$. Because this pair does not breed during the second month, $f_2 = 1$ also. To find the number of pairs after n months, add the number on the island the previous month, f_{n-1}, and the number of newborn pairs, which equals f_{n-2}, because each newborn pair comes from a pair at least 2 months old.

The Fibonacci numbers appear in many other places in nature, including the number of petals on flowers and the number of spirals on seedheads.

Consequently, the sequence $\{f_n\}$ satisfies the recurrence relation

$$f_n = f_{n-1} + f_{n-2}$$

for $n \geq 3$ together with the initial conditions $f_1 = 1$ and $f_2 = 1$. Because this recurrence relation and the initial conditions uniquely determine this sequence, the number of pairs of rabbits on the island after n months is given by the nth Fibonacci number. ◀

Demo

Example 2 involves a famous puzzle.

EXAMPLE 2 **The Tower of Hanoi Puzzle** A popular puzzle of the late nineteenth century invented by the French mathematician Édouard Lucas, called the Tower of Hanoi, consists of three pegs mounted on a board together with disks of different sizes. Initially these disks are placed on the first peg in order of size, with the largest on the bottom (as shown in Figure 2). The rules of the puzzle allow disks to be moved one at a time from one peg to another as long as a disk is never placed on top of a smaller disk. The goal of the puzzle is to have all the disks on the second peg in order of size, with the largest on the bottom.

Links

Let H_n denote the number of moves needed to solve the Tower of Hanoi puzzle with n disks. Set up a recurrence relation for the sequence $\{H_n\}$.

Solution: Begin with n disks on peg 1. We can transfer the top $n - 1$ disks, following the rules of the puzzle, to peg 3 using H_{n-1} moves (see Figure 3 for an illustration of the pegs and disks at this point). We keep the largest disk fixed during these moves. Then, we use one move to transfer the largest disk to the second peg. Finally, we transfer the $n - 1$ disks on peg 3 to peg 2 using H_{n-1} moves, placing them on top of the largest disk, which always stays fixed on the bottom of peg 2. This shows that we can solve the Tower of Hano puzzle for n disks using $2H_{n-1} + 1$ moves.

Schemes for efficiently backing up computer files on multiple tapes or other media are based on the moves used to solve the Tower of Hanoi puzzle.

We now show that we cannot solve the puzzle for n disks using fewer that $2H_{n-1} + 1$ moves. Note that when we move the largest disk, we must have already moved the $n - 1$ smaller disks onto a peg other than peg 1. Doing so requires at least H_{n-1} moves. Another move is needed to

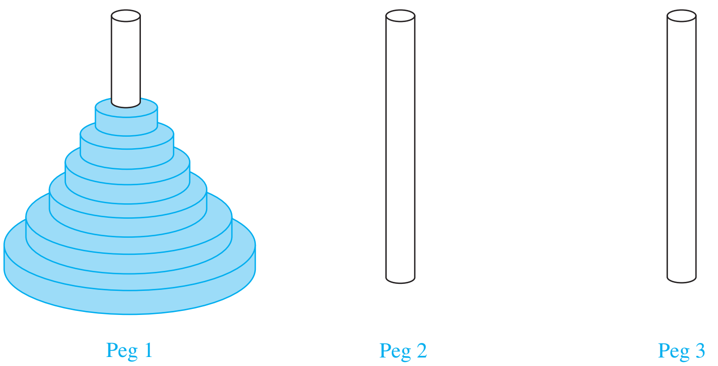


FIGURE 2 The initial position in the Tower of Hanoi.

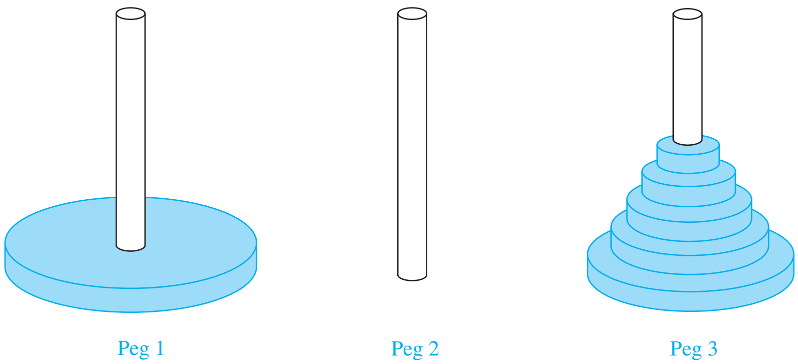


FIGURE 3 **An intermediate position in the Tower of Hanoi.**

transfer the largest disk. Finally, at least H_{n-1} more moves are needed to put the $n-1$ smallest disks back on top of the largest disk. Adding the number of moves required gives us the desired lower bound.

We conclude that

$$H_n = 2H_{n-1} + 1.$$

The initial condition is $H_1 = 1$, because one disk can be transferred from peg 1 to peg 2, according to the rules of the puzzle, in one move.

We can use an iterative approach to solve this recurrence relation. Note that

$$\begin{aligned}
H_n &= 2H_{n-1} + 1 \\
&= 2(2H_{n-2} + 1) + 1 = 2^2H_{n-2} + 2 + 1 \\
&= 2^2(2H_{n-3} + 1) + 2 + 1 = 2^3H_{n-3} + 2^2 + 2 + 1 \\
&\;\vdots \\
&= 2^{n-1}H_1 + 2^{n-2} + 2^{n-3} + \cdots + 2 + 1 \\
&= 2^{n-1} + 2^{n-2} + \cdots + 2 + 1 \\
&= 2^n - 1.
\end{aligned}$$

We have used the recurrence relation repeatedly to express H_n in terms of previous terms of the sequence. In the next to last equality, the initial condition $H_1 = 1$ has been used. The last equality is based on the formula for the sum of the terms of a geometric series, which can be found in Theorem 1 in Section 2.4.

The iterative approach has produced the solution to the recurrence relation $H_n = 2H_{n-1} + 1$ with the initial condition $H_1 = 1$. This formula can be proved using mathematical induction. This is left for the reader as Exercise 1.

A myth created to accompany the puzzle tells of a tower in Hanoi where monks are transferring 64 gold disks from one peg to another, according to the rules of the puzzle. The myth says that the world will end when they finish the puzzle. How long after the monks started will the world end if the monks take one second to move a disk?

From the explicit formula, the monks require

$$2^{64} - 1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$$

moves to transfer the disks. Making one move per second, it will take them more than 500 billion years to complete the transfer, so the world should survive a while longer than it already has. ◀

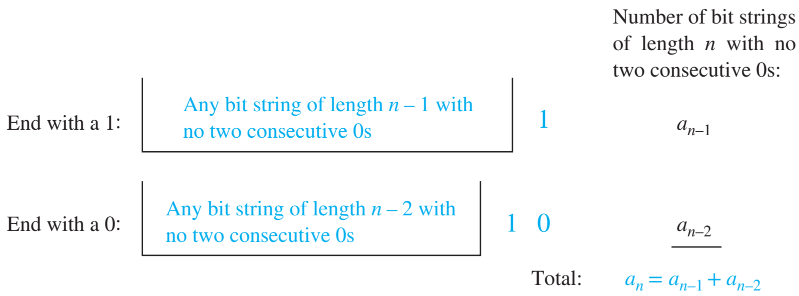


FIGURE 4 **Counting bit strings of length n with no two consecutive 0s.**

Links

Remark: Many people have studied variations of the original Tower of Hanoi puzzle discussed in Example 2. Some variations use more pegs, some allow disks to be of the same size, and some restrict the types of allowable disk moves. One of the oldest and most interesting variations is the **Reve's puzzle**,* proposed in 1907 by Henry Dudeney in his book *The Canterbury Puzzles*. The Reve's puzzle involves pilgrims challenged by the Reve to move a stack of cheese wheels of varying sizes from the first of four stools to another stool without ever placing a cheese wheel on one of smaller diameter. The Reve's puzzle, expressed in terms of pegs and disks, follows the same rules as the Tower of Hanoi puzzle, except that four pegs are used. Similarly, we can generalize the Tower of Hanoi puzzle where there are p pegs, where p is an integer greater than three. You may find it surprising that no one has been able to establish the minimum number of moves required to solve the generalization of this puzzle for p pegs. (Note that there have been some published claims that this problem has been solved, but these are not accepted by experts.) However, in 2014 Thierry Bousch showed that the minimum number of moves required when there are four pegs equals the number of moves used by an algorithm invented by Frame and Stewart in 1939. (See Exercises 38–45 and [St94] and [Bo14] for more information.)

Example 3 illustrates how recurrence relations can be used to count bit strings of a specified length that have a certain property.

EXAMPLE 3 Find a recurrence relation and give initial conditions for the number of bit strings of length n that do not have two consecutive 0s. How many such bit strings are there of length five?

Solution: Let a_n denote the number of bit strings of length n that do not have two consecutive 0s. We assume that $n \geq 3$, so that the bit string has at least three bits. Strings of this sort of length n can be divided into those that end in 1 and those that end in 0. The bit strings of length n ending with 1 that do not have two consecutive 0s are precisely the bit strings of length $n-1$ with no two consecutive 0s with a 1 added at the end. Consequently, there are a_{n-1} such bit strings.

Bit strings of length n ending with a 0 that do not have two consecutive 0s must have 1 as their $(n-1)$st bit; otherwise they would end with a pair of 0s. Hence, the bit strings of length n ending with a 0 that have no two consecutive 0s are precisely the bit strings of length $n-2$ with no two consecutive 0s with 10 added at the end. Consequently, there are a_{n-2} such bit strings.

We conclude, as illustrated in Figure 4, that

$$a_n = a_{n-1} + a_{n-2}$$

for $n \geq 3$.

***Reve*, more commonly spelled *reeve*, is an archaic word for *governor*.

The initial conditions are $a_1 = 2$, because both bit strings of length one, 0 and 1 do not have consecutive 0s, and $a_2 = 3$, because the valid bit strings of length two are 01, 10, and 11. To obtain a_5, we use the recurrence relation three times to find that

$$
\begin{aligned}
a_3 &= a_2 + a_1 = 3 + 2 = 5,\\
a_4 &= a_3 + a_2 = 5 + 3 = 8,\\
a_5 &= a_4 + a_3 = 8 + 5 = 13.
\end{aligned}
$$

◀

Remark: Note that $\{a_n\}$ satisfies the same recurrence relation as the Fibonacci sequence. Because $a_1 = f_3$ and $a_2 = f_4$ it follows that $a_n = f_{n+2}$.

Example 4 shows how a recurrence relation can be used to model the number of codewords that are allowable using certain validity checks.

EXAMPLE 4 **Codeword Enumeration** A computer system considers a string of decimal digits a valid codeword if it contains an even number of 0 digits. For instance, 1230407869 is valid, whereas 120987045608 is not valid. Let a_n be the number of valid n-digit codewords. Find a recurrence relation for a_n.

Solution: Note that $a_1 = 9$ because there are 10 one-digit strings, and only one, namely, the string 0, is not valid. A recurrence relation can be derived for this sequence by considering how a valid n-digit string can be obtained from strings of $n - 1$ digits. There are two ways to form a valid string with n digits from a string with one fewer digit.

First, a valid string of n digits can be obtained by appending a valid string of $n - 1$ digits with a digit other than 0. This appending can be done in nine ways. Hence, a valid string with n digits can be formed in this manner in $9a_{n-1}$ ways.

Second, a valid string of n digits can be obtained by appending a 0 to a string of length $n - 1$ that is not valid. (This produces a string with an even number of 0 digits because the invalid string of length $n - 1$ has an odd number of 0 digits.) The number of ways that this can be done equals the number of invalid $(n - 1)$-digit strings. Because there are 10^{n-1} strings of length $n - 1$, and a_{n-1} are valid, there are $10^{n-1} - a_{n-1}$ valid n-digit strings obtained by appending an invalid string of length $n - 1$ with a 0.

Because all valid strings of length n are produced in one of these two ways, it follows that there are

$$
\begin{aligned}
a_n &= 9a_{n-1} + (10^{n-1} - a_{n-1})\\
&= 8a_{n-1} + 10^{n-1}
\end{aligned}
$$

valid strings of length n. ◀

Example 5 establishes a recurrence relation that appears in many different contexts.

EXAMPLE 5 Find a recurrence relation for C_n, the number of ways to parenthesize the product of $n + 1$ numbers, $x_0 \cdot x_1 \cdot x_2 \cdot \cdots \cdot x_n$, to specify the order of multiplication. For example, $C_3 = 5$ because there are five ways to parenthesize $x_0 \cdot x_1 \cdot x_2 \cdot x_3$ to determine the order of multiplication:

$$
\begin{array}{lll}
((x_0 \cdot x_1) \cdot x_2) \cdot x_3 & (x_0 \cdot (x_1 \cdot x_2)) \cdot x_3 & (x_0 \cdot x_1) \cdot (x_2 \cdot x_3)\\
x_0 \cdot ((x_1 \cdot x_2) \cdot x_3) & x_0 \cdot (x_1 \cdot (x_2 \cdot x_3)).
\end{array}
$$

Solution: To develop a recurrence relation for C_n, we note that however we insert parentheses in the product $x_0 \cdot x_1 \cdot x_2 \cdot \cdots \cdot x_n$, one "$\cdot$" operator remains outside all parentheses, namely, the operator for the final multiplication to be performed. [For example, in $(x_0 \cdot (x_1 \cdot x_2)) \cdot x_3$, it is the final "$\cdot$", while in $(x_0 \cdot x_1) \cdot (x_2 \cdot x_3)$ it is the second "$\cdot$".] This final operator appears between two of the $n + 1$ numbers, say, x_k and x_{k+1}. There are $C_k C_{n-k-1}$ ways to insert parentheses to determine the order of the $n + 1$ numbers to be multiplied when the final operator appears between x_k and x_{k+1}, because there are C_k ways to insert parentheses in the product $x_0 \cdot x_1 \cdot \cdots \cdot x_k$ to determine the order in which these $k + 1$ numbers are to be multiplied and C_{n-k-1} ways to insert parentheses in the product $x_{k+1} \cdot x_{k+2} \cdot \cdots \cdot x_n$ to determine the order in which these $n - k$ numbers are to be multiplied. Because this final operator can appear between any two of the $n + 1$ numbers, it follows that

$$\begin{aligned} C_n &= C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-2} C_1 + C_{n-1} C_0 \\ &= \sum_{k=0}^{n-1} C_k C_{n-k-1}. \end{aligned}$$

Note that the initial conditions are $C_0 = 1$ and $C_1 = 1$. ◀

The recurrence relation in Example 5 can be solved using the method of generating functions, which will be discussed in Section 8.4. It can be shown that $C_n = C(2n, n)/(n + 1)$ (see Exercise 43 in Section 8.4) and that $C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ (see [GrKnPa94]). The sequence $\{C_n\}$ is the sequence of **Catalan numbers**, named after Eugène Charles Catalan. This sequence appears as the solution of many different counting problems besides the one considered here (see the chapter on Catalan numbers in [MiRo91] or [RoTe03] for details).

Links

8.1.3 Algorithms and Recurrence Relations

Recurrence relations play an important role in many aspects of the study of algorithms and their complexity. In Section 8.3, we will show how recurrence relations can be used to analyze the complexity of divide-and-conquer algorithms, such as the merge sort algorithm introduced in Section 5.4. As we will see in Section 8.3, divide-and-conquer algorithms recursively divide a problem into a fixed number of nonoverlapping subproblems until they become simple enough to solve directly. We conclude this section by introducing another algorithmic paradigm known as **dynamic programming**, which can be used to solve many optimization problems efficiently.

An algorithm follows the dynamic programming paradigm when it recursively breaks down a problem into simpler overlapping subproblems, and computes the solution using the solutions of the subproblems. Generally, recurrence relations are used to find the overall solution from the solutions of the subproblems. Dynamic programming has been used to solve important problems in such diverse areas as economics, computer vision, speech recognition, artificial intelligence, computer graphics, and bioinformatics. In this section we will illustrate the use of dynamic programming by constructing an algorithm for solving a scheduling problem. Before doing so, we will relate the amusing origin of the name *dynamic programming*, which was introduced by the mathematician Richard Bellman in the 1950s. Bellman was working at the RAND Corporation on projects for the U.S. military, and at that time, the U.S. Secretary of Defense was hostile to mathematical research. Bellman decided that to ensure funding, he needed a name not containing the word mathematics for his method for solving scheduling and planning problems. He decided to use the adjective *dynamic* because, as he said "it's impossible to use the word dynamic in a pejorative sense" and he thought that dynamic programming was "something not even a Congressman could object to."

Links

AN EXAMPLE OF DYNAMIC PROGRAMMING The problem we use to illustrate dynamic programming is related to the problem studied in Example 7 in Section 3.1. In that problem our goal was to schedule as many talks as possible in a single lecture hall. These talks have preset start and end times; once a talk starts, it continues until it ends; no two talks can proceed at the same time; and a talk can begin at the same time another one ends. We developed a greedy algorithm that always produces an optimal schedule, as we proved in Example 12 in Section 5.1. Now suppose that our goal is not to schedule the most talks possible, but rather to have the largest possible combined attendance of the scheduled talks.

We formalize this problem by supposing that we have n talks, where talk j begins at time t_j, ends at time e_j, and will be attended by w_j students. We want a schedule that maximizes the total number of student attendees. That is, we wish to schedule a subset of talks to maximize the sum of w_j over all scheduled talks. (Note that when a student attends more than one talk, this student is counted according to the number of talks attended.) We denote by $T(j)$ the maximum number of total attendees for an optimal schedule from the first j talks, so $T(n)$ is the maximal number of total attendees for an optimal schedule for all n talks.

We first sort the talks in order of increasing end time. After doing this, we renumber the talks so that $e_1 \leq e_2 \leq \cdots \leq e_n$. We say that two talks are **compatible** if they can be part of the same schedule, that is, if the times they are scheduled do not overlap (other than the possibility one ends and the other starts at the same time). We define $p(j)$ to be largest integer i, $i < j$, for which $e_i \leq s_j$, if such an integer exists, and $p(j) = 0$ otherwise. That is, talk $p(j)$ is the talk ending latest among talks compatible with talk j that end before talk j ends, if such a talk exists, and $p(j) = 0$ if there are no such talks.

EXAMPLE 6 Consider seven talks with these start times and end times, as illustrated in Figure 5.

Talk 1: start 8 A.M., end 10 A.M.
Talk 2: start 9 A.M., end 11 A.M.
Talk 3: start 10:30 A.M., end 12 noon
Talk 4: start 9:30 A.M., end 1 P.M.
Talk 5: start 8:30 A.M., end 2 P.M.
Talk 6: start 11 A.M., end 2 P.M.
Talk 7: start 1 P.M., end 2 P.M.

Find $p(j)$ for $j = 1, 2, \ldots, 7$.

Solution: We have $p(1) = 0$ and $p(2) = 0$, because no talks end before either of the first two talks begin. We have $p(3) = 1$ because talk 3 and talk 1 are compatible, but talk 3 and talk 2 are not compatible; $p(4) = 0$ because talk 4 is not compatible with any of talks 1, 2, and 3; $p(5) = 0$

Links

©Paul Fearn/Alamy Stock Photo

EUGÈNE CHARLES CATALAN (1814–1894) Eugène Catalan was born in Bruges, then part of France. His father became a successful architect in Paris while Eugène was a boy. Catalan attended a Parisian school for design hoping to follow in his father's footsteps. At 15, he won the job of teaching geometry to his design school classmates. After graduating, Catalan attended a school for the fine arts, but because of his mathematical aptitude his instructors recommended that he enter the École Polytechnique. He became a student there, but after his first year, he was expelled because of his politics. However, he was readmitted, and in 1835, he graduated and won a position at the Collège de Châlons sur Marne.

In 1838, Catalan returned to Paris where he founded a preparatory school with two other mathematicians, Sturm and Liouville. After teaching there for a short time, he was appointed to a position at the École Polytechnique. He received his doctorate from the École Polytechnique in 1841, but his political activity in favor of the French Republic hurt his career prospects. In 1846 Catalan held a position at the Collège de Charlemagne; he was appointed to the Lycée Saint Louis in 1849. However, when Catalan would not take a required oath of allegiance to the new Emperor Louis-Napoleon Bonaparte, he lost his job. For 13 years he held no permanent position. Finally, in 1865 he was appointed to a chair of mathematics at the University of Liège, Belgium, a position he held until his 1884 retirement.

Catalan made many contributions to number theory and to the related subject of continued fractions. He defined what are now known as the Catalan numbers when he solved the problem of dissecting a polygon into triangles using non-intersecting diagonals. Catalan is also well known for formulating what was known as the *Catalan conjecture*. This asserted that 8 and 9 are the only consecutive powers of integers, a conjecture not solved until 2003. Catalan wrote many textbooks, including several that became quite popular and appeared in as many as 12 editions. Perhaps this textbook will have a 12th edition someday!

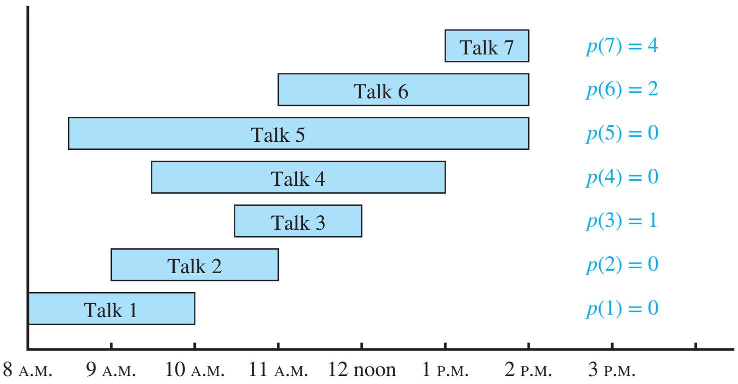


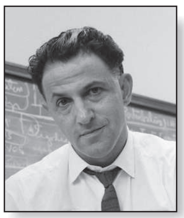
FIGURE 5 **A schedule of lectures with the values of $p(n)$ shown.**

because talk 5 is not compatible with any of talks 1, 2, 3, and 4; and $p(6) = 2$ because talk 6 and talk 2 are compatible, but talk 6 is not compatible with any of talks 3, 4, and 5. Finally, $p(7) = 4$, because talk 7 and talk 4 are compatible, but talk 7 is not compatible with either of talks 5 or 6. ◀

To develop a dynamic programming algorithm for this problem, we first develop a key recurrence relation. To do this, first note that if $j \leq n$, there are two possibilities for an optimal schedule of the first j talks (recall that we are assuming that the n talks are ordered by increasing end time): *(i)* talk j belongs to the optimal schedule or *(ii)* it does not.

Case (i): We know that talks $p(j) + 1, \ldots, j - 1$ do not belong to this schedule, for none of these other talks are compatible with talk j. Furthermore, the other talks in this optimal schedule must comprise an optimal schedule for talks $1, 2, \ldots, p(j)$. For if there were a better schedule for talks

Links



©Alfred Eisenstaedt/The LIFE Picture Collection/Getty Images

RICHARD BELLMAN (1920–1984) Richard Bellman, born in Brooklyn, where his father was a grocer, spent many hours in the museums and libraries of New York as a child. After graduating high school, he studied mathematics at Brooklyn College and graduated in 1941. He began postgraduate work at Johns Hopkins University, but because of the war, left to teach electronics at the University of Wisconsin. He was able to continue his mathematics studies at Wisconsin, and in 1943 he received his masters degree there. Later, Bellman entered Princeton University, teaching in a special U.S. Army program. In late 1944, he was drafted into the army. He was assigned to the Manhattan Project at Los Alamos where he worked in theoretical physics. After the war, he returned to Princeton and received his Ph.D. in 1946.

After briefly teaching at Princeton, he moved to Stanford University, where he attained tenure. At Stanford he pursued his fascination with number theory. However, Bellman decided to focus on mathematical questions arising from real-world problems. In 1952, he joined the RAND Corporation, working on multistage decision processes, operations research problems, and applications to the social sciences and medicine. He worked on many military projects while at RAND. In 1965 he left RAND to become professor of mathematics, electrical and biomedical engineering and medicine at the University of Southern California.

In the 1950s Bellman pioneered the use of dynamic programming, a technique invented earlier, in a wide range of settings. He is also known for his work on stochastic control processes, in which he introduced what is now called the Bellman equation. He coined the term *curse of dimensionality* to describe problems caused by the exponential increase in volume associated with adding extra dimensions to a space. He wrote an amazing number of books and research papers with many coauthors, including many on industrial production and economic systems. His work led to the application of computing techniques in a wide variety of areas ranging from the design of guidance systems for space vehicles, to network optimization, and even to pest control.

Tragically, in 1973 Bellman was diagnosed with a brain tumor. Although it was removed successfully, complications left him severely disabled. Fortunately, he managed to continue his research and writing during his remaining ten years of life. Bellman received many prizes and awards, including the first Norbert Wiener Prize in Applied Mathematics and the IEEE Gold Medal of Honor. He was elected to the National Academy of Sciences. He was held in high regard for his achievements, courage, and admirable qualities. Bellman was the father of two children.

$1, 2, \ldots, p(j)$, by adding talk j, we will have a schedule better than the overall optimal schedule. Consequently, in case *(i)*, we have $T(j) = w_j + T(p(j))$.

Case (ii): When talk j does not belong to an optimal schedule, it follows that an optimal schedule from talks $1, 2, \ldots, j$ is the same as an optimal schedule from talks $1, 2, \ldots, j-1$. Consequently, in case *(ii)*, we have $T(j) = T(j-1)$. Combining cases *(i)* and *(ii)* leads us to the recurrence relation

$$T(j) = \max(w_j + T(p(j)), T(j-1)).$$

Now that we have developed this recurrence relation, we can construct an efficient algorithm, Algorithm 1, for computing the maximum total number of attendees. We ensure that the algorithm is efficient by storing the value of each $T(j)$ after we compute it. This allows us to compute $T(j)$ only once. If we did not do this, the algorithm would have exponential worst-case complexity. The process of storing the values as each is computed is known as **memoization** and is an important technique for making recursive algorithms efficient.

ALGORITHM 1 Dynamic Programming Algorithm for Scheduling Talks.

procedure *Maximum Attendees* ($s_1, s_2, \ldots, s_n$: start times of talks;
  $e_1, e_2, \ldots, e_n$: end times of talks; $w_1, w_2, \ldots, w_n$: number of attendees to talks)
  sort talks by end time and relabel so that $e_1 \le e_2 \le \cdots \le e_n$
for $j := 1$ **to** n
    if no job i with $i < j$ is compatible with job j
        $p(j) = 0$
    else $p(j) := \max\{i — i < j$ and job i is compatible with job $j\}$
    $T(0) := 0$
for $j := 1$ **to** n
    $T(j) := \max(w_j + T(p(j)), T(j-1))$
return $T(n)\{T(n)$ is the maximum number of attendees$\}$

In Algorithm 1 we determine the maximum number of attendees that can be achieved by a schedule of talks, but we do not find a schedule that achieves this maximum. To find talks we need to schedule, we use the fact that talk j belongs to an optimal solution for the first j talks if and only if $w_j + T(p(j)) \ge T(j-1)$. We leave it as Exercise 53 to construct an algorithm based on this observation that determines which talks should be scheduled to achieve the maximum total number of attendees.

Algorithm 1 is a good example of dynamic programming as the maximum total attendance is found using the optimal solutions of the overlapping subproblems, each of which determines the maximum total attendance of the first j talks for some j with $1 \le j \le n-1$. See Exercises 56 and 57 and Supplementary Exercises 14 and 17 for other examples of dynamic programming.

Exercises

1. Use mathematical induction to verify the formula derived in Example 2 for the number of moves required to complete the Tower of Hanoi puzzle.

2. a) Find a recurrence relation for the number of permutations of a set with n elements.

b) Use this recurrence relation to find the number of permutations of a set with n elements using iteration.

3. A vending machine dispensing books of stamps accepts only one-dollar coins, \$1 bills, and \$5 bills.

a) Find a recurrence relation for the number of ways to deposit n dollars in the vending machine, where the order in which the coins and bills are deposited matters.

b) What are the initial conditions?

c) How many ways are there to deposit $10 for a book of stamps?

4. A country uses as currency coins with values of 1 peso, 2 pesos, 5 pesos, and 10 pesos and bills with values of 5 pesos, 10 pesos, 20 pesos, 50 pesos, and 100 pesos. Find a recurrence relation for the number of ways to pay a bill of n pesos if the order in which the coins and bills are paid matters.

5. How many ways are there to pay a bill of 17 pesos using the currency described in Exercise 4, where the order in which coins and bills are paid matters?

***6. a)** Find a recurrence relation for the number of strictly increasing sequences of positive integers that have 1 as their first term and n as their last term, where n is a positive integer. That is, sequences $a_1, a_2, \ldots, a_k$, where $a_1 = 1$, $a_k = n$, and $a_j < a_{j+1}$ for $j = 1, 2, \ldots, k-1$.

b) What are the initial conditions?

c) How many sequences of the type described in (a) are there when n is an integer with $n \geq 2$?

7. a) Find a recurrence relation for the number of bit strings of length n that contain a pair of consecutive 0s.

b) What are the initial conditions?

c) How many bit strings of length seven contain two consecutive 0s?

8. a) Find a recurrence relation for the number of bit strings of length n that contain three consecutive 0s.

b) What are the initial conditions?

c) How many bit strings of length seven contain three consecutive 0s?

9. a) Find a recurrence relation for the number of bit strings of length n that do not contain three consecutive 0s.

b) What are the initial conditions?

c) How many bit strings of length seven do not contain three consecutive 0s?

***10. a)** Find a recurrence relation for the number of bit strings of length n that contain the string 01.

b) What are the initial conditions?

c) How many bit strings of length seven contain the string 01?

11. a) Find a recurrence relation for the number of ways to climb n stairs if the person climbing the stairs can take one stair or two stairs at a time.

b) What are the initial conditions?

c) In how many ways can this person climb a flight of eight stairs?

12. a) Find a recurrence relation for the number of ways to climb n stairs if the person climbing the stairs can take one, two, or three stairs at a time.

b) What are the initial conditions?

c) In how many ways can this person climb a flight of eight stairs?

A string that contains only 0s, 1s, and 2s is called a **ternary string**.

13. a) Find a recurrence relation for the number of ternary strings of length n that do not contain two consecutive 0s.

b) What are the initial conditions?

c) How many ternary strings of length six do not contain two consecutive 0s?

14. a) Find a recurrence relation for the number of ternary strings of length n that contain two consecutive 0s.

b) What are the initial conditions?

c) How many ternary strings of length six contain two consecutive 0s?

***15. a)** Find a recurrence relation for the number of ternary strings of length n that do not contain two consecutive 0s or two consecutive 1s.

b) What are the initial conditions?

c) How many ternary strings of length six do not contain two consecutive 0s or two consecutive 1s?

***16. a)** Find a recurrence relation for the number of ternary strings of length n that contain either two consecutive 0s or two consecutive 1s.

b) What are the initial conditions?

c) How many ternary strings of length six contain two consecutive 0s or two consecutive 1s?

***17. a)** Find a recurrence relation for the number of ternary strings of length n that do not contain consecutive symbols that are the same.

b) What are the initial conditions?

c) How many ternary strings of length six do not contain consecutive symbols that are the same?

****18. a)** Find a recurrence relation for the number of ternary strings of length n that contain two consecutive symbols that are the same.

b) What are the initial conditions?

c) How many ternary strings of length six contain consecutive symbols that are the same?

19. Messages are transmitted over a communications channel using two signals. The transmittal of one signal requires 1 microsecond, and the transmittal of the other signal requires 2 microseconds.

a) Find a recurrence relation for the number of different messages consisting of sequences of these two signals, where each signal in the message is immediately followed by the next signal, that can be sent in n microseconds.

b) What are the initial conditions?

c) How many different messages can be sent in 10 microseconds using these two signals?

20. A bus driver pays all tolls, using only nickels and dimes, by throwing one coin at a time into the mechanical toll collector.

a) Find a recurrence relation for the number of different ways the bus driver can pay a toll of n cents (where the order in which the coins are used matters).

b) In how many different ways can the driver pay a toll of 45 cents?

21. a) Find the recurrence relation satisfied by R_n, where R_n is the number of regions that a plane is divided into by n lines, if no two of the lines are parallel and no three of the lines go through the same point.

b) Find R_n using iteration.

*22. a) Find the recurrence relation satisfied by R_n, where R_n is the number of regions into which the surface of a sphere is divided by n great circles (which are the intersections of the sphere and planes passing through the center of the sphere), if no three of the great circles go through the same point.

b) Find R_n using iteration.

*23. a) Find the recurrence relation satisfied by S_n, where S_n is the number of regions into which three-dimensional space is divided by n planes if every three of the planes meet in one point, but no four of the planes go through the same point.

b) Find S_n using iteration.

24. Find a recurrence relation for the number of bit sequences of length n with an even number of 0s.

25. How many bit sequences of length seven contain an even number of 0s?

26. a) Find a recurrence relation for the number of ways to completely cover a $2 \times n$ checkerboard with 1×2 dominoes. [*Hint:* Consider separately the coverings where the position in the top right corner of the checkerboard is covered by a domino positioned horizontally and where it is covered by a domino positioned vertically.]

b) What are the initial conditions for the recurrence relation in part (a)?

c) How many ways are there to completely cover a 2×17 checkerboard with 1×2 dominoes?

27. a) Find a recurrence relation for the number of ways to lay out a walkway with slate tiles if the tiles are red, green, or gray, so that no two red tiles are adjacent and tiles of the same color are considered indistinguishable.

b) What are the initial conditions for the recurrence relation in part (a)?

c) How many ways are there to lay out a path of seven tiles as described in part (a)?

28. Show that the Fibonacci numbers satisfy the recurrence relation $f_n = 5f_{n-4} + 3f_{n-5}$ for $n = 5, 6, 7, \ldots,$ together with the initial conditions $f_0 = 0, f_1 = 1, f_2 = 1, f_3 = 2,$ and $f_4 = 3$. Use this recurrence relation to show that f_{5n} is divisible by 5, for $n = 1, 2, 3, \ldots$.

*29. Let $S(m, n)$ denote the number of onto functions from a set with m elements to a set with n elements. Show that $S(m, n)$ satisfies the recurrence relation

$$S(m, n) = n^m - \sum_{k=1}^{n-1} C(n, k)S(m, k)$$

whenever $m \geq n$ and $n > 1$, with the initial condition $S(m, 1) = 1$.

30. a) Write out all the ways the product $x_0 \cdot x_1 \cdot x_2 \cdot x_3 \cdot x_4$ can be parenthesized to determine the order of multiplication.

b) Use the recurrence relation developed in Example 5 to calculate C_4, the number of ways to parenthesize the product of five numbers so as to determine the order of multiplication. Verify that you listed the correct number of ways in part (a).

c) Check your result in part (b) by finding C_4, using the closed formula for C_n mentioned in the solution of Example 5.

31. a) Use the recurrence relation developed in Example 5 to determine C_5, the number of ways to parenthesize the product of six numbers so as to determine the order of multiplication.

b) Check your result with the closed formula for C_5 mentioned in the solution of Example 5.

*32. In the Tower of Hanoi puzzle, suppose our goal is to transfer all n disks from peg 1 to peg 3, but we cannot move a disk directly between pegs 1 and 3. Each move of a disk must be a move involving peg 2. As usual, we cannot place a disk on top of a smaller disk.

a) Find a recurrence relation for the number of moves required to solve the puzzle for n disks with this added restriction.

b) Solve this recurrence relation to find a formula for the number of moves required to solve the puzzle for n disks.

c) How many different arrangements are there of the n disks on three pegs so that no disk is on top of a smaller disk?

d) Show that every allowable arrangement of the n disks occurs in the solution of this variation of the puzzle.

Exercises 33–37 deal with a variation of the **Josephus problem** described by Graham, Knuth, and Patashnik in [GrKnPa94]. This problem is based on an account by the historian Flavius Josephus, who was part of a band of 41 Jewish rebels trapped in a cave by the Romans during the Jewish-Roman war of the first century. The rebels preferred suicide to capture; they decided to form a circle and to repeatedly count off around the circle, killing every third rebel left alive. However, Josephus and another rebel did not want to be killed this way; they determined the positions where they should stand to be the last two rebels remaining alive. The variation we consider begins with n people, numbered 1 to n, standing around a circle. In each stage, every second person still left alive is eliminated until only one survives. We denote the number of the survivor by $J(n)$.

33. Determine the value of $J(n)$ for each integer n with $1 \leq n \leq 16$.

34. Use the values you found in Exercise 33 to conjecture a formula for $J(n)$. [*Hint:* Write $n = 2^m + k$, where m is a nonnegative integer and k is a nonnegative integer less than 2^m.]

35. Show that $J(n)$ satisfies the recurrence relation $J(2n) = 2J(n) - 1$ and $J(2n + 1) = 2J(n) + 1$, for $n \geq 1$, and $J(1) = 1$.

36. Use mathematical induction to prove the formula you conjectured in Exercise 34, making use of the recurrence relation from Exercise 35.

37. Determine $J(100)$, $J(1000)$, and $J(10{,}000)$ from your formula for $J(n)$.

Exercises 38–45 involve the Reve's puzzle, the variation of the Tower of Hanoi puzzle with four pegs and n disks. Before presenting these exercises, we describe the Frame–Stewart algorithm for moving the disks from peg 1 to peg 4 so that no disk is ever on top of a smaller one. This algorithm, given the number of disks n as input, depends on a choice of an integer k with $1 \leq k \leq n$. When there is only one disk, move it from peg 1 to peg 4 and stop. For $n > 1$, the algorithm proceeds recursively, using these three steps. Recursively move the stack of the $n - k$ smallest disks from peg 1 to peg 2, using all four pegs. Next move the stack of the k largest disks from peg 1 to peg 4, using the three-peg algorithm from the Tower of Hanoi puzzle without using the peg holding the $n - k$ smallest disks. Finally, recursively move the smallest $n - k$ disks to peg 4, using all four pegs. Frame and Stewart showed that to produce the fewest moves using their algorithm, k should be chosen to be the smallest integer such that n does not exceed $t_k = k(k+1)/2$, the kth triangular number, that is, $t_{k-1} < n \leq t_k$. The long-standing conjecture, known as **Frame's conjecture**, that this algorithm uses the fewest number of moves required to solve the puzzle, was proved by Thierry Bousch in 2014.

38. Show that the Reve's puzzle with three disks can be solved using five, and no fewer, moves.

39. Show that the Reve's puzzle with four disks can be solved using nine, and no fewer, moves.

40. Describe the moves made by the Frame–Stewart algorithm, with k chosen so that the fewest moves are required, for

a) 5 disks. **b)** 6 disks. **c)** 7 disks. **d)** 8 disks.

***41.** Show that if $R(n)$ is the number of moves used by the Frame–Stewart algorithm to solve the Reve's puzzle with n disks, where k is chosen to be the smallest integer with $n \leq k(k+1)/2$, then $R(n)$ satisfies the recurrence relation $R(n) = 2R(n-k) + 2^k - 1$, with $R(0) = 0$ and $R(1) = 1$.

***42.** Show that if k is as chosen in Exercise 41, then $R(n) - R(n-1) = 2^{k-1}$.

***43.** Show that if k is as chosen in Exercise 41, then $R(n) = \sum_{i=1}^{k} i2^{i-1} - (t_k - n)2^{k-1}$.

***44.** Use Exercise 43 to give an upper bound on the number of moves required to solve the Reve's puzzle for all integers n with $1 \leq n \leq 25$.

***45.** Show that $R(n)$ is $O(\sqrt{n}2^{\sqrt{2n}})$.

Let $\{a_n\}$ be a sequence of real numbers. The **backward differences** of this sequence are defined recursively as shown next. The **first difference** ∇a_n is

$$\nabla a_n = a_n - a_{n-1}.$$

The **$(k+1)$st difference** $\nabla^{k+1} a_n$ is obtained from $\nabla^k a_n$ by

$$\nabla^{k+1} a_n = \nabla^k a_n - \nabla^k a_{n-1}.$$

46. Find ∇a_n for the sequence $\{a_n\}$, where

a) $a_n = 4$. **b)** $a_n = 2n$.

c) $a_n = n^2$. **d)** $a_n = 2^n$.

47. Find $\nabla^2 a_n$ for the sequences in Exercise 46.

48. Show that $a_{n-1} = a_n - \nabla a_n$.

49. Show that $a_{n-2} = a_n - 2\nabla a_n + \nabla^2 a_n$.

***50.** Prove that a_{n-k} can be expressed in terms of a_n, ∇a_n, $\nabla^2 a_n, \ldots, \nabla^k a_n$.

51. Express the recurrence relation $a_n = a_{n-1} + a_{n-2}$ in terms of a_n, ∇a_n, and $\nabla^2 a_n$.

52. Show that any recurrence relation for the sequence $\{a_n\}$ can be written in terms of a_n, ∇a_n, $\nabla^2 a_n, \ldots$. The resulting equation involving the sequences and its differences is called a **difference equation**.

***53.** Construct the algorithm described in the text after Algorithm 1 for determining which talks should be scheduled to maximize the total number of attendees and not just the maximum total number of attendees determined by Algorithm 1.

54. Use Algorithm 1 to determine the maximum number of total attendees in the talks in Example 6 if w_i, the number of attendees of talk i, $i = 1, 2, \ldots, 7$, is

a) 20, 10, 50, 30, 15, 25, 40.

b) 100, 5, 10, 20, 25, 40, 30.

c) 2, 3, 8, 5, 4, 7, 10.

d) 10, 8, 7, 25, 20, 30, 5.

55. For each part of Exercise 54, use your algorithm from Exercise 53 to find the optimal schedule for talks so that the total number of attendees is maximized.

56. In this exercise we will develop a dynamic programming algorithm for finding the maximum sum of consecutive terms of a sequence of real numbers. That is, given a sequence of real numbers $a_1, a_2, \ldots, a_n$, the algorithm computes the maximum sum $\sum_{i=j}^{k} a_i$ where $1 \leq j \leq k \leq n$.

a) Show that if all terms of the sequence are nonnegative, this problem is solved by taking the sum of all terms. Then, give an example where the maximum sum of consecutive terms is not the sum of all terms.

b) Let $M(k)$ be the maximum of the sums of consecutive terms of the sequence ending at a_k. That is, $M(k) = \max_{1 \leq j \leq k} \sum_{i=j}^{k} a_i$. Explain why the recurrence relation $M(k) = \max(M(k-1) + a_k, a_k)$ holds for $k = 2, \ldots, n$.

c) Use part (b) to develop a dynamic programming algorithm for solving this problem.

d) Show each step your algorithm from part (c) uses to find the maximum sum of consecutive terms of the sequence 2, −3, 4, 1, −2, 3.

e) Show that the worst-case complexity in terms of the number of additions and comparisons of your algorithm from part (c) is linear.

*57. Dynamic programming can be used to develop an algorithm for solving the matrix-chain multiplication problem introduced in Section 3.3. This is the problem of determining how the product $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_n$ can be computed using the fewest integer multiplications, where $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ are $m_1 \times m_2, m_2 \times m_3, \ldots, m_n \times m_{n+1}$ matrices, respectively, and each matrix has integer entries. Recall that by the associative law, the product does not depend on the order in which the matrices are multiplied.

a) Show that the brute-force method of determining the minimum number of integer multiplications needed to solve a matrix-chain multiplication problem has exponential worst-case complexity. [*Hint:* Do this by first showing that the order of multiplication of matrices is specified by parenthesizing the product. Then, use Example 5 and the result of part (c) of Exercise 43 in Section 8.4.]

b) Denote by $\mathbf{A}_{ij}$ the product $\mathbf{A}_i\mathbf{A}_{i+1} \ldots, \mathbf{A}_j$, and $M(i, j)$ the minimum number of integer multiplications required to find $\mathbf{A}_{ij}$. Show that if the least number of integer multiplications are used to compute $\mathbf{A}_{ij}$, where $i < j$, by splitting the product into the product of $\mathbf{A}_i$ through $\mathbf{A}_k$ and the product of $\mathbf{A}_{k+1}$ through $\mathbf{A}_j$, then the first k terms must be parenthesized so that $\mathbf{A}_{ik}$ is computed in the optimal way using $M(i, k)$ integer multiplications, and $\mathbf{A}_{k+1,j}$ must be parenthesized so that $\mathbf{A}_{k+1,j}$ is computed in the optimal way using $M(k + 1, j)$ integer multiplications.

c) Explain why part (b) leads to the recurrence relation $M(i, j) = \min_{i \le k < j}(M(i, k) + M(k + 1, j) + m_i m_{k+1} m_{j+1})$ if $1 \le i \le j < j \le n$.

d) Use the recurrence relation in part (c) to construct an efficient algorithm for determining the order the n matrices should be multiplied to use the minimum number of integer multiplications. Store the partial results $M(i, j)$ as you find them so that your algorithm will not have exponential complexity.

e) Show that your algorithm from part (d) has $O(n^3)$ worst-case complexity in terms of multiplications of integers.

8.2 Solving Linear Recurrence Relations

8.2.1 Introduction

Links

A wide variety of recurrence relations occur in models. Some of these recurrence relations can be solved using iteration or some other ad hoc technique. However, one important class of recurrence relations can be explicitly solved in a systematic way. These are recurrence relations that express the terms of a sequence as linear combinations of previous terms.

Definition 1

A *linear homogeneous recurrence relation of degree k with constant coefficients* is a recurrence relation of the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k},$$

where $c_1, c_2, \ldots, c_k$ are real numbers, and $c_k \neq 0$.

The recurrence relation in the definition is **linear** because the right-hand side is a sum of previous terms of the sequence each multiplied by a function of n. The recurrence relation is **homogeneous** because no terms occur that are not multiples of the a_js. The coefficients of the terms of the sequence are all **constants**, rather than functions that depend on n. The **degree** is k because a_n is expressed in terms of the previous k terms of the sequence.

A consequence of the second principle of mathematical induction is that a sequence satisfying the recurrence relation in the definition is uniquely determined by this recurrence relation and the k initial conditions

$$a_0 = C_0, a_1 = C_1, \ldots, a_{k-1} = C_{k-1}.$$

EXAMPLE 1 The recurrence relation $P_n = (1.11)P_{n-1}$ is a linear homogeneous recurrence relation of degree one. The recurrence relation $f_n = f_{n-1} + f_{n-2}$ is a linear homogeneous recurrence relation of degree two. The recurrence relation $a_n = a_{n-5}$ is a linear homogeneous recurrence relation of degree five. ◀

To help clarify the definition of linear homogeneous recurrence relations with constant coefficients, we will now provide examples of recurrence relations each lacking one of the defining properties.

EXAMPLE 2 The recurrence relation $a_n = a_{n-1} + a_{n-2}^2$ is not linear. The recurrence relation $H_n = 2H_{n-1} + 1$ is not homogeneous. The recurrence relation $B_n = nB_{n-1}$ does not have constant coefficients. ◀

Linear homogeneous recurrence relations are studied for two reasons. First, they often occur in modeling of problems. Second, they can be systematically solved.

8.2.2 Solving Linear Homogeneous Recurrence Relations with Constant Coefficients

Recurrence relations may be difficult to solve, but fortunately this is not the case for linear homogenous recurrence relations with constant coefficients. We can use two key ideas to find all their solutions. First, these recurrence relations have solutions of the form $a_n = r^n$, where r is a constant. To see this, observe that $a_n = r^n$ is a solution of the recurrence relation $a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k}$ if and only if

$$r^n = c_1r^{n-1} + c_2r^{n-2} + \cdots + c_kr^{n-k}.$$

When both sides of this equation are divided by r^{n-k} (when $r \neq 0$) and the right-hand side is subtracted from the left, we obtain the equation

$$r^k - c_1r^{k-1} - c_2r^{k-2} - \cdots - c_{k-1}r - c_k = 0.$$

Consequently, the sequence $\{a_n\}$ with $a_n = r^n$ where $r \neq 0$ is a solution if and only if r is a solution of this last equation. We call this the **characteristic equation** of the recurrence relation. The solutions of this equation are called the **characteristic roots** of the recurrence relation. As we will see, these characteristic roots can be used to give an explicit formula for all the solutions of the recurrence relation.

The other key observation is that a linear combination of two solutions of a linear homogeneous recurrence relation is also a solution. To see this, suppose that s_n and t_n are both solutions of $a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k}$. Then

$$s_n = c_1s_{n-1} + c_2s_{n-2} + \cdots + c_ks_{n-k}$$

and

$$t_n = c_1t_{n-1} + c_2t_{n-2} + \cdots + c_kt_{n-k}.$$

Now suppose that b_1 and b_2 are real numbers. Then

$$\begin{aligned} b_1s_n + b_2t_n &= b_1(c_1s_{n-1} + c_2s_{n-2} + \cdots + c_ks_{n-k}) + b_2(c_1t_{n-1} + c_2t_{n-2} + \cdots + c_kt_{n-k}) \\ &= c_1(b_1s_{n-1} + b_2t_{n-1}) + c_2(b_1s_{n-2} + b_2t_{n-2}) + \cdots + c_k(b_1s_{n-k} + b_kt_{n-k}). \end{aligned}$$

This means that $b_1s_n + b_2t_n$ is also a solution of the same linear homogeneous recurrence relation.

Using these key observations, we will show how to solve linear homogeneous recurrence relations with constant coefficients.

THE DEGREE TWO CASE We now turn our attention to linear homogeneous recurrence relations of degree two. First, consider the case when there are two distinct characteristic roots.

THEOREM 1 Let c_1 and c_2 be real numbers. Suppose that $r^2 - c_1 r - c_2 = 0$ has two distinct roots r_1 and r_2. Then the sequence $\{a_n\}$ is a solution of the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ if and only if $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ for $n = 0, 1, 2, \ldots$, where α_1 and α_2 are constants.

Proof: We must do two things to prove the theorem. First, it must be shown that if r_1 and r_2 are the roots of the characteristic equation, and α_1 and α_2 are constants, then the sequence $\{a_n\}$ with $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ is a solution of the recurrence relation. Second, it must be shown that if the sequence $\{a_n\}$ is a solution, then $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ for some constants α_1 and α_2.

We now show that if $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$, then the sequence $\{a_n\}$ is a solution of the recurrence relation. Because r_1 and r_2 are roots of $r^2 - c_1 r - c_2 = 0$, it follows that $r_1^2 = c_1 r_1 + c_2$ and $r_2^2 = c_1 r_2 + c_2$.

From these equations, we see that

$$\begin{aligned} c_1 a_{n-1} + c_2 a_{n-2} &= c_1(\alpha_1 r_1^{n-1} + \alpha_2 r_2^{n-1}) + c_2(\alpha_1 r_1^{n-2} + \alpha_2 r_2^{n-2}) \\ &= \alpha_1 r_1^{n-2}(c_1 r_1 + c_2) + \alpha_2 r_2^{n-2}(c_1 r_2 + c_2) \\ &= \alpha_1 r_1^{n-2} r_1^2 + \alpha_2 r_2^{n-2} r_2^2 \\ &= \alpha_1 r_1^n + \alpha_2 r_2^n \\ &= a_n. \end{aligned}$$

This shows that the sequence $\{a_n\}$ with $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ is a solution of the recurrence relation.

To show that every solution $\{a_n\}$ of the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ has $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ for $n = 0, 1, 2, \ldots$, for some constants α_1 and α_2, suppose that $\{a_n\}$ is a solution of the recurrence relation, and the initial conditions $a_0 = C_0$ and $a_1 = C_1$ hold. It will be shown that there are constants α_1 and α_2 such that the sequence $\{a_n\}$ with $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ satisfies these same initial conditions. This requires that

$$\begin{aligned} a_0 &= C_0 = \alpha_1 + \alpha_2, \\ a_1 &= C_1 = \alpha_1 r_1 + \alpha_2 r_2. \end{aligned}$$

We can solve these two equations for α_1 and α_2. From the first equation it follows that $\alpha_2 = C_0 - \alpha_1$. Inserting this expression into the second equation gives

$$C_1 = \alpha_1 r_1 + (C_0 - \alpha_1) r_2.$$

Hence,

$$C_1 = \alpha_1(r_1 - r_2) + C_0 r_2.$$

This shows that

$$\alpha_1 = \frac{C_1 - C_0 r_2}{r_1 - r_2}$$

and

$$\alpha_2 = C_0 - \alpha_1 = C_0 - \frac{C_1 - C_0 r_2}{r_1 - r_2} = \frac{C_0 r_1 - C_1}{r_1 - r_2},$$

where these expressions for α_1 and α_2 depend on the fact that $r_1 \neq r_2$. (When $r_1 = r_2$, this theorem is not true.) Hence, with these values for α_1 and α_2, the sequence $\{a_n\}$ with $\alpha_1 r_1^n + \alpha_2 r_2^n$ satisfies the two initial conditions.

We know that $\{a_n\}$ and $\{\alpha_1 r_1^n + \alpha_2 r_2^n\}$ are both solutions of the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ and both satisfy the initial conditions when $n = 0$ and $n = 1$. Because there is a unique solution of a linear homogeneous recurrence relation of degree two with two initial conditions, it follows that the two solutions are the same, that is, $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$ for all nonnegative integers n. We have completed the proof by showing that a solution of the linear homogeneous recurrence relation with constant coefficients of degree two must be of the form $a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$, where α_1 and α_2 are constants. ◁

The characteristic roots of a linear homogeneous recurrence relation with constant coefficients may be complex numbers. Theorem 1 (and also subsequent theorems in this section) still applies in this case. Recurrence relations with complex characteristic roots will not be discussed in the text. Readers familiar with complex numbers may wish to solve Exercises 38 and 39.

Examples 3 and 4 show how to use Theorem 1 to solve recurrence relations.

EXAMPLE 3 What is the solution of the recurrence relation

Extra Examples

$$a_n = a_{n-1} + 2a_{n-2}$$

with $a_0 = 2$ and $a_1 = 7$?

Solution: Theorem 1 can be used to solve this problem. The characteristic equation of the recurrence relation is $r^2 - r - 2 = 0$. Its roots are $r = 2$ and $r = -1$. Hence, the sequence $\{a_n\}$ is a solution to the recurrence relation if and only if

$$a_n = \alpha_1 2^n + \alpha_2 (-1)^n,$$

for some constants α_1 and α_2. From the initial conditions, it follows that

$$a_0 = 2 = \alpha_1 + \alpha_2,$$
$$a_1 = 7 = \alpha_1 \cdot 2 + \alpha_2 \cdot (-1).$$

Solving these two equations shows that $\alpha_1 = 3$ and $\alpha_2 = -1$. Hence, the solution to the recurrence relation and initial conditions is the sequence $\{a_n\}$ with

$$a_n = 3 \cdot 2^n - (-1)^n.$$ ◀

EXAMPLE 4 Find an explicit formula for the Fibonacci numbers.

Solution: Recall that the sequence of Fibonacci numbers satisfies the recurrence relation $f_n = f_{n-1} + f_{n-2}$ and also satisfies the initial conditions $f_0 = 0$ and $f_1 = 1$. The roots of the characteristic equation $r^2 - r - 1 = 0$ are $r_1 = (1 + \sqrt{5})/2$ and $r_2 = (1 - \sqrt{5})/2$. Therefore, from Theorem 1 it follows that the Fibonacci numbers are given by

$$f_n = \alpha_1 \left(\frac{1 + \sqrt{5}}{2}\right)^n + \alpha_2 \left(\frac{1 - \sqrt{5}}{2}\right)^n,$$

for some constants α_1 and α_2. The initial conditions $f_0 = 0$ and $f_1 = 1$ can be used to find these constants. We have

$$f_0 = \alpha_1 + \alpha_2 = 0,$$
$$f_1 = \alpha_1 \left(\frac{1+\sqrt{5}}{2}\right) + \alpha_2 \left(\frac{1-\sqrt{5}}{2}\right) = 1.$$

The solution to these simultaneous equations for α_1 and α_2 is

$$\alpha_1 = 1/\sqrt{5}, \qquad \alpha_2 = -1/\sqrt{5}.$$

Consequently, the Fibonacci numbers are given by

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

◀

Theorem 1 does not apply when there is one characteristic root of multiplicity two. If this happens, then $a_n = nr_0^n$ is another solution of the recurrence relation when r_0 is a root of multiplicity two of the characteristic equation. Theorem 2 shows how to handle this case.

THEOREM 2 Let c_1 and c_2 be real numbers with $c_2 \neq 0$. Suppose that $r^2 - c_1 r - c_2 = 0$ has only one root r_0. A sequence $\{a_n\}$ is a solution of the recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ if and only if $a_n = \alpha_1 r_0^n + \alpha_2 n r_0^n$, for $n = 0, 1, 2, \ldots$, where α_1 and α_2 are constants.

The proof of Theorem 2 is left as Exercise 10. Example 5 illustrates the use of this theorem.

EXAMPLE 5 What is the solution of the recurrence relation

$$a_n = 6a_{n-1} - 9a_{n-2}$$

with initial conditions $a_0 = 1$ and $a_1 = 6$?

Solution: The only root of $r^2 - 6r + 9 = 0$ is $r = 3$. Hence, the solution to this recurrence relation is

$$a_n = \alpha_1 3^n + \alpha_2 n 3^n$$

for some constants α_1 and α_2. Using the initial conditions, it follows that

$$a_0 = 1 = \alpha_1,$$
$$a_1 = 6 = \alpha_1 \cdot 3 + \alpha_2 \cdot 3.$$

Solving these two equations shows that $\alpha_1 = 1$ and $\alpha_2 = 1$. Consequently, the solution to this recurrence relation and the initial conditions is

$$a_n = 3^n + n3^n.$$

◀

THE GENERAL CASE We will now state the general result about the solution of linear homogeneous recurrence relations with constant coefficients, where the degree may be greater than two, under the assumption that the characteristic equation has distinct roots. The proof of this result will be left as Exercise 16.

THEOREM 3 Let $c_1, c_2, \ldots, c_k$ be real numbers. Suppose that the characteristic equation

$$r^k - c_1 r^{k-1} - \cdots - c_k = 0$$

has k distinct roots $r_1, r_2, \ldots, r_k$. Then a sequence $\{a_n\}$ is a solution of the recurrence relation

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k}$$

if and only if

$$a_n = \alpha_1 r_1^n + \alpha_2 r_2^n + \cdots + \alpha_k r_k^n$$

for $n = 0, 1, 2, \ldots$, where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are constants.

We illustrate the use of the theorem with Example 6.

EXAMPLE 6 Find the solution to the recurrence relation

$$a_n = 6a_{n-1} - 11a_{n-2} + 6a_{n-3}$$

with the initial conditions $a_0 = 2$, $a_1 = 5$, and $a_2 = 15$.

Solution: The characteristic polynomial of this recurrence relation is

$$r^3 - 6r^2 + 11r - 6.$$

The characteristic roots are $r = 1$, $r = 2$, and $r = 3$, because $r^3 - 6r^2 + 11r - 6 = (r-1)(r-2)(r-3)$. Hence, the solutions to this recurrence relation are of the form

$$a_n = \alpha_1 \cdot 1^n + \alpha_2 \cdot 2^n + \alpha_3 \cdot 3^n.$$

To find the constants α_1, α_2, and α_3, use the initial conditions. This gives

$$\begin{aligned} a_0 &= 2 = \alpha_1 + \alpha_2 + \alpha_3, \\ a_1 &= 5 = \alpha_1 + \alpha_2 \cdot 2 + \alpha_3 \cdot 3, \\ a_2 &= 15 = \alpha_1 + \alpha_2 \cdot 4 + \alpha_3 \cdot 9. \end{aligned}$$

When these three simultaneous equations are solved for α_1, α_2, and α_3, we find that $\alpha_1 = 1$, $\alpha_2 = -1$, and $\alpha_3 = 2$. Hence, the unique solution to this recurrence relation and the given initial conditions is the sequence $\{a_n\}$ with

$$a_n = 1 - 2^n + 2 \cdot 3^n.$$ ◀

We now state the most general result about linear homogeneous recurrence relations with constant coefficients, allowing the characteristic equation to have multiple roots. The key point is that for each root r of the characteristic equation, the general solution has a summand of the form $P(n)r^n$, where $P(n)$ is a polynomial of degree $m - 1$, with m the multiplicity of this root. We leave the proof of this result as Exercise 51.

THEOREM 4 Let $c_1, c_2, \ldots, c_k$ be real numbers. Suppose that the characteristic equation

$$r^k - c_1 r^{k-1} - \cdots - c_k = 0$$

has t distinct roots $r_1, r_2, \ldots, r_t$ with multiplicities $m_1, m_2, \ldots, m_t$, respectively, so that $m_i \geq 1$ for $i = 1, 2, \ldots, t$ and $m_1 + m_2 + \cdots + m_t = k$. Then a sequence $\{a_n\}$ is a solution of the recurrence relation

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k}$$

if and only if

$$\begin{aligned} a_n = {} & (\alpha_{1,0} + \alpha_{1,1} n + \cdots + \alpha_{1,m_1-1} n^{m_1-1}) r_1^n \\ & + (\alpha_{2,0} + \alpha_{2,1} n + \cdots + \alpha_{2,m_2-1} n^{m_2-1}) r_2^n \\ & + \cdots + (\alpha_{t,0} + \alpha_{t,1} n + \cdots + \alpha_{t,m_t-1} n^{m_t-1}) r_t^n \end{aligned}$$

for $n = 0, 1, 2, \ldots$, where $\alpha_{i,j}$ are constants for $1 \leq i \leq t$ and $0 \leq j \leq m_i - 1$.

Example 7 illustrates how Theorem 4 is used to find the general form of a solution of a linear homogeneous recurrence relation when the characteristic equation has several repeated roots.

EXAMPLE 7 Suppose that the roots of the characteristic equation of a linear homogeneous recurrence relation are 2, 2, 2, 5, 5, and 9 (that is, there are three roots, the root 2 with multiplicity three, the root 5 with multiplicity two, and the root 9 with multiplicity one). What is the form of the general solution?

Solution: By Theorem 4, the general form of the solution is

$$(\alpha_{1,0} + \alpha_{1,1} n + \alpha_{1,2} n^2) 2^n + (\alpha_{2,0} + \alpha_{2,1} n) 5^n + \alpha_{3,0} 9^n.$$

◀

We now illustrate the use of Theorem 4 to solve a linear homogeneous recurrence relation with constant coefficients when the characteristic equation has a root of multiplicity three.

EXAMPLE 8 Find the solution to the recurrence relation

$$a_n = -3a_{n-1} - 3a_{n-2} - a_{n-3}$$

with initial conditions $a_0 = 1$, $a_1 = -2$, and $a_2 = -1$.

Solution: The characteristic equation of this recurrence relation is

$$r^3 + 3r^2 + 3r + 1 = 0.$$

Because $r^3 + 3r^2 + 3r + 1 = (r + 1)^3$, there is a single root $r = -1$ of multiplicity three of the characteristic equation. By Theorem 4 the solutions of this recurrence relation are of the form

$$a_n = \alpha_{1,0}(-1)^n + \alpha_{1,1} n(-1)^n + \alpha_{1,2} n^2 (-1)^n.$$

To find the constants $\alpha_{1,0}$, $\alpha_{1,1}$, and $\alpha_{1,2}$, use the initial conditions. This gives

$$a_0 = 1 = \alpha_{1,0},$$
$$a_1 = -2 = -\alpha_{1,0} - \alpha_{1,1} - \alpha_{1,2},$$
$$a_2 = -1 = \alpha_{1,0} + 2\alpha_{1,1} + 4\alpha_{1,2}.$$

The simultaneous solution of these three equations is $\alpha_{1,0} = 1$, $\alpha_{1,1} = 3$, and $\alpha_{1,2} = -2$. Hence, the unique solution to this recurrence relation and the given initial conditions is the sequence $\{a_n\}$ with

$$a_n = (1 + 3n - 2n^2)(-1)^n.$$

◀

8.2.3 Linear Nonhomogeneous Recurrence Relations with Constant Coefficients

We have seen how to solve linear homogeneous recurrence relations with constant coefficients. Is there a relatively simple technique for solving a linear, but not homogeneous, recurrence relation with constant coefficients, such as $a_n = 3a_{n-1} + 2n$? We will see that the answer is yes for certain families of such recurrence relations.

The recurrence relation $a_n = 3a_{n-1} + 2n$ is an example of a **linear nonhomogeneous recurrence relation with constant coefficients**, that is, a recurrence relation of the form

$$a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k} + F(n),$$

where $c_1, c_2, \ldots, c_k$ are real numbers and $F(n)$ is a function not identically zero depending only on n. The recurrence relation

$$a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k}$$

is called the **associated homogeneous recurrence relation**. It plays an important role in the solution of the nonhomogeneous recurrence relation.

EXAMPLE 9 Each of the recurrence relations $a_n = a_{n-1} + 2^n$, $a_n = a_{n-1} + a_{n-2} + n^2 + n + 1$, $a_n = 3a_{n-1} + n3^n$, and $a_n = a_{n-1} + a_{n-2} + a_{n-3} + n!$ is a linear nonhomogeneous recurrence relation with constant coefficients. The associated linear homogeneous recurrence relations are $a_n = a_{n-1}$, $a_n = a_{n-1} + a_{n-2}$, $a_n = 3a_{n-1}$, and $a_n = a_{n-1} + a_{n-2} + a_{n-3}$, respectively. ◀

The key fact about linear nonhomogeneous recurrence relations with constant coefficients is that every solution is the sum of a particular solution and a solution of the associated linear homogeneous recurrence relation, as Theorem 5 shows.

THEOREM 5 If $\{a_n^{(p)}\}$ is a particular solution of the nonhomogeneous linear recurrence relation with constant coefficients

$$a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k} + F(n),$$

then every solution is of the form $\{a_n^{(p)} + a_n^{(h)}\}$, where $\{a_n^{(h)}\}$ is a solution of the associated homogeneous recurrence relation

$$a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_ka_{n-k}.$$

Proof: Because $\{a_n^{(p)}\}$ is a particular solution of the nonhomogeneous recurrence relation, we know that

$$a_n^{(p)} = c_1 a_{n-1}^{(p)} + c_2 a_{n-2}^{(p)} + \cdots + c_k a_{n-k}^{(p)} + F(n).$$

Now suppose that $\{b_n\}$ is a second solution of the nonhomogeneous recurrence relation, so that

$$b_n = c_1 b_{n-1} + c_2 b_{n-2} + \cdots + c_k b_{n-k} + F(n).$$

Subtracting the first of these two equations from the second shows that

$$b_n - a_n^{(p)} = c_1(b_{n-1} - a_{n-1}^{(p)}) + c_2(b_{n-2} - a_{n-2}^{(p)}) + \cdots + c_k(b_{n-k} - a_{n-k}^{(p)}).$$

It follows that $\{b_n - a_n^p\}$ is a solution of the associated homogeneous linear recurrence, say, $\{a_n^{(h)}\}$. Consequently, $b_n = a_n^{(p)} + a_n^{(h)}$ for all n. ◁

By Theorem 5, we see that the key to solving nonhomogeneous recurrence relations with constant coefficients is finding a particular solution. Then every solution is a sum of this solution and a solution of the associated homogeneous recurrence relation. Although there is no general method for finding such a solution that works for every function $F(n)$, there are techniques that work for certain types of functions $F(n)$, such as polynomials and powers of constants. This is illustrated in Examples 10 and 11.

EXAMPLE 10 Find all solutions of the recurrence relation $a_n = 3a_{n-1} + 2n$. What is the solution with $a_1 = 3$?

Solution: To solve this linear nonhomogeneous recurrence relation with constant coefficients, we need to solve its associated linear homogeneous equation and to find a particular solution for the given nonhomogeneous equation. The associated linear homogeneous equation is $a_n = 3a_{n-1}$. Its solutions are $a_n^{(h)} = \alpha 3^n$, where α is a constant.

We now find a particular solution. Because $F(n) = 2n$ is a polynomial in n of degree one, a reasonable trial solution is a linear function in n, say, $p_n = cn + d$, where c and d are constants. To determine whether there are any solutions of this form, suppose that $p_n = cn + d$ is such a solution. Then the equation $a_n = 3a_{n-1} + 2n$ becomes $cn + d = 3(c(n-1) + d) + 2n$. Simplifying and combining like terms gives $(2 + 2c)n + (2d - 3c) = 0$. It follows that $cn + d$ is a solution if and only if $2 + 2c = 0$ and $2d - 3c = 0$. This shows that $cn + d$ is a solution if and only if $c = -1$ and $d = -3/2$. Consequently, $a_n^{(p)} = -n - 3/2$ is a particular solution.

By Theorem 5 all solutions are of the form

$$a_n = a_n^{(p)} + a_n^{(h)} = -n - \frac{3}{2} + \alpha \cdot 3^n,$$

where α is a constant.

To find the solution with $a_1 = 3$, let $n = 1$ in the formula we obtained for the general solution. We find that $3 = -1 - 3/2 + 3\alpha$, which implies that $\alpha = 11/6$. The solution we seek is $a_n = -n - 3/2 + (11/6)3^n$. ◀

EXAMPLE 11 Find all solutions of the recurrence relation

Extra Examples

$$a_n = 5a_{n-1} - 6a_{n-2} + 7^n.$$

Solution: This is a linear nonhomogeneous recurrence relation. The solutions of its associated homogeneous recurrence relation

$$a_n = 5a_{n-1} - 6a_{n-2}$$

are $a_n^{(h)} = \alpha_1 \cdot 3^n + \alpha_2 \cdot 2^n$, where α_1 and α_2 are constants. Because $F(n) = 7^n$, a reasonable trial solution is $a_n^{(p)} = C \cdot 7^n$, where C is a constant. Substituting the terms of this sequence into the recurrence relation implies that $C \cdot 7^n = 5C \cdot 7^{n-1} - 6C \cdot 7^{n-2} + 7^n$. Factoring out 7^{n-2}, this equation becomes $49C = 35C - 6C + 49$, which implies that $20C = 49$, or that $C = 49/20$. Hence, $a_n^{(p)} = (49/20)7^n$ is a particular solution. By Theorem 5, all solutions are of the form

$$a_n = \alpha_1 \cdot 3^n + \alpha_2 \cdot 2^n + (49/20)7^n.$$ ◀

In Examples 10 and 11, we made an educated guess that there are solutions of a particular form. In both cases we were able to find particular solutions. This was not an accident. Whenever $F(n)$ is the product of a polynomial in n and the nth power of a constant, we know exactly what form a particular solution has, as stated in Theorem 6. We leave the proof of Theorem 6 as Exercise 52.

THEOREM 6 Suppose that $\{a_n\}$ satisfies the linear nonhomogeneous recurrence relation

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_k a_{n-k} + F(n),$$

where $c_1, c_2, \ldots, c_k$ are real numbers, and

$$F(n) = (b_t n^t + b_{t-1} n^{t-1} + \cdots + b_1 n + b_0)s^n,$$

where $b_0, b_1, \ldots, b_t$ and s are real numbers. When s is not a root of the characteristic equation of the associated linear homogeneous recurrence relation, there is a particular solution of the form

$$(p_t n^t + p_{t-1} n^{t-1} + \cdots + p_1 n + p_0)s^n.$$

When s is a root of this characteristic equation and its multiplicity is m, there is a particular solution of the form

$$n^m(p_t n^t + p_{t-1} n^{t-1} + \cdots + p_1 n + p_0)s^n.$$

Note that in the case when s is a root of multiplicity m of the characteristic equation of the associated linear homogeneous recurrence relation, the factor n^m ensures that the proposed particular solution will not already be a solution of the associated linear homogeneous recurrence relation. We next provide Example 12 to illustrate the form of a particular solution provided by Theorem 6.

EXAMPLE 12 What form does a particular solution of the linear nonhomogeneous recurrence relation $a_n = 6a_{n-1} - 9a_{n-2} + F(n)$ have when $F(n) = 3^n$, $F(n) = n3^n$, $F(n) = n^2 2^n$, and $F(n) = (n^2 + 1)3^n$?

Solution: The associated linear homogeneous recurrence relation is $a_n = 6a_{n-1} - 9a_{n-2}$. Its characteristic equation, $r^2 - 6r + 9 = (r - 3)^2 = 0$, has a single root, 3, of multiplicity two. To apply Theorem 6, with $F(n)$ of the form $P(n)s^n$, where $P(n)$ is a polynomial and s is a constant, we need to ask whether s is a root of this characteristic equation.

Because $s = 3$ is a root with multiplicity $m = 2$ but $s = 2$ is not a root, Theorem 6 tells us that a particular solution has the form $p_0 n^2 3^n$ if $F(n) = 3^n$, the form $n^2(p_1 n + p_0)3^n$ if $F(n) =$

$n3^n$, the form $(p_2n^2 + p_1n + p_0)2^n$ if $F(n) = n^2 2^n$, and the form $n^2(p_2n^2 + p_1n + p_0)3^n$ if $F(n) = (n^2 + 1)3^n$. ◀

Care must be taken when $s = 1$ when solving recurrence relations of the type covered by Theorem 6. In particular, to apply this theorem with $F(n) = b_t n_t + b_{t-1} n_{t-1} + \cdots + b_1 n + b_0$, the parameter s takes the value $s = 1$ (even though the term 1^n does not explicitly appear). By the theorem, the form of the solution then depends on whether 1 is a root of the characteristic equation of the associated linear homogeneous recurrence relation. This is illustrated in Example 13, which shows how Theorem 6 can be used to find a formula for the sum of the first n positive integers.

EXAMPLE 13 Let a_n be the sum of the first n positive integers, so that

$$a_n = \sum_{k=1}^{n} k.$$

Note that a_n satisfies the linear nonhomogeneous recurrence relation

$$a_n = a_{n-1} + n.$$

(To obtain a_n, the sum of the first n positive integers, from a_{n-1}, the sum of the first $n - 1$ positive integers, we add n.) Note that the initial condition is $a_1 = 1$.

The associated linear homogeneous recurrence relation for a_n is

$$a_n = a_{n-1}.$$

The solutions of this homogeneous recurrence relation are given by $a_n^{(h)} = c(1)^n = c$, where c is a constant. To find all solutions of $a_n = a_{n-1} + n$, we need find only a single particular solution. By Theorem 6, because $F(n) = n = n \cdot (1)^n$ and $s = 1$ is a root of degree one of the characteristic equation of the associated linear homogeneous recurrence relation, there is a particular solution of the form $n(p_1n + p_0) = p_1n^2 + p_0n$.

Inserting this into the recurrence relation gives $p_1n^2 + p_0n = p_1(n-1)^2 + p_0(n-1) + n$. Simplifying, we see that $n(2p_1 - 1) + (p_0 - p_1) = 0$, which means that $2p_1 - 1 = 0$ and $p_0 - p_1 = 0$, so $p_0 = p_1 = 1/2$. Hence,

$$a_n^{(p)} = \frac{n^2}{2} + \frac{n}{2} = \frac{n(n+1)}{2}$$

is a particular solution. Hence, all solutions of the original recurrence relation $a_n = a_{n-1} + n$ are given by $a_n = a_n^{(h)} + a_n^{(p)} = c + n(n+1)/2$. Because $a_1 = 1$, we have $1 = a_1 = c + 1 \cdot 2/2 = c + 1$, so $c = 0$. It follows that $a_n = n(n+1)/2$. (This is the same formula given in Table 2 in Section 2.4 and derived previously.) ◀

Exercises

1. Determine which of these are linear homogeneous recurrence relations with constant coefficients. Also, find the degree of those that are.

a) $a_n = 3a_{n-1} + 4a_{n-2} + 5a_{n-3}$
b) $a_n = 2na_{n-1} + a_{n-2}$ **c)** $a_n = a_{n-1} + a_{n-4}$
d) $a_n = a_{n-1} + 2$ **e)** $a_n = a_{n-1}^2 + a_{n-2}$
f) $a_n = a_{n-2}$ **g)** $a_n = a_{n-1} + n$

2. Determine which of these are linear homogeneous recurrence relations with constant coefficients. Also, find the degree of those that are.

a) $a_n = 3a_{n-2}$ **b)** $a_n = 3$
c) $a_n = a_{n-1}^2$ **d)** $a_n = a_{n-1} + 2a_{n-3}$
e) $a_n = a_{n-1}/n$
f) $a_n = a_{n-1} + a_{n-2} + n + 3$
g) $a_n = 4a_{n-2} + 5a_{n-4} + 9a_{n-7}$

3. Solve these recurrence relations together with the initial conditions given.

a) $a_n = 2a_{n-1}$ for $n \geq 1$, $a_0 = 3$
b) $a_n = a_{n-1}$ for $n \geq 1$, $a_0 = 2$
c) $a_n = 5a_{n-1} - 6a_{n-2}$ for $n \geq 2$, $a_0 = 1$, $a_1 = 0$
d) $a_n = 4a_{n-1} - 4a_{n-2}$ for $n \geq 2$, $a_0 = 6$, $a_1 = 8$
e) $a_n = -4a_{n-1} - 4a_{n-2}$ for $n \geq 2$, $a_0 = 0$, $a_1 = 1$
f) $a_n = 4a_{n-2}$ for $n \geq 2$, $a_0 = 0$, $a_1 = 4$
g) $a_n = a_{n-2}/4$ for $n \geq 2$, $a_0 = 1$, $a_1 = 0$

4. Solve these recurrence relations together with the initial conditions given.

a) $a_n = a_{n-1} + 6a_{n-2}$ for $n \geq 2$, $a_0 = 3$, $a_1 = 6$
b) $a_n = 7a_{n-1} - 10a_{n-2}$ for $n \geq 2$, $a_0 = 2$, $a_1 = 1$
c) $a_n = 6a_{n-1} - 8a_{n-2}$ for $n \geq 2$, $a_0 = 4$, $a_1 = 10$
d) $a_n = 2a_{n-1} - a_{n-2}$ for $n \geq 2$, $a_0 = 4$, $a_1 = 1$
e) $a_n = a_{n-2}$ for $n \geq 2$, $a_0 = 5$, $a_1 = -1$
f) $a_n = -6a_{n-1} - 9a_{n-2}$ for $n \geq 2$, $a_0 = 3$, $a_1 = -3$
g) $a_{n+2} = -4a_{n+1} + 5a_n$ for $n \geq 0$, $a_0 = 2$, $a_1 = 8$

5. How many different messages can be transmitted in n microseconds using the two signals described in Exercise 19 in Section 8.1?

6. How many different messages can be transmitted in n microseconds using three different signals if one signal requires 1 microsecond for transmittal, the other two signals require 2 microseconds each for transmittal, and a signal in a message is followed immediately by the next signal?

7. In how many ways can a $2 \times n$ rectangular checkerboard be tiled using 1×2 and 2×2 pieces?

8. A model for the number of lobsters caught per year is based on the assumption that the number of lobsters caught in a year is the average of the number caught in the two previous years.

a) Find a recurrence relation for $\{L_n\}$, where L_n is the number of lobsters caught in year n, under the assumption for this model.

b) Find L_n if 100,000 lobsters were caught in year 1 and 300,000 were caught in year 2.

9. A deposit of $100,000 is made to an investment fund at the beginning of a year. On the last day of each year two dividends are awarded. The first dividend is 20% of the amount in the account during that year. The second dividend is 45% of the amount in the account in the previous year.

a) Find a recurrence relation for $\{P_n\}$, where P_n is the amount in the account at the end of n years if no money is ever withdrawn.

b) How much is in the account after n years if no money has been withdrawn?

***10.** Prove Theorem 2.

11. The **Lucas numbers** satisfy the recurrence relation

Links

$$L_n = L_{n-1} + L_{n-2},$$

and the initial conditions $L_0 = 2$ and $L_1 = 1$.

a) Show that $L_n = f_{n-1} + f_{n+1}$ for $n = 2, 3, \ldots$, where f_n is the nth Fibonacci number.

b) Find an explicit formula for the Lucas numbers.

12. Find the solution to $a_n = 2a_{n-1} + a_{n-2} - 2a_{n-3}$ for $n = 3, 4, 5, \ldots$, with $a_0 = 3$, $a_1 = 6$, and $a_2 = 0$.

13. Find the solution to $a_n = 7a_{n-2} + 6a_{n-3}$ with $a_0 = 9$, $a_1 = 10$, and $a_2 = 32$.

14. Find the solution to $a_n = 5a_{n-2} - 4a_{n-4}$ with $a_0 = 3$, $a_1 = 2$, $a_2 = 6$, and $a_3 = 8$.

15. Find the solution to $a_n = 2a_{n-1} + 5a_{n-2} - 6a_{n-3}$ with $a_0 = 7$, $a_1 = -4$, and $a_2 = 8$.

***16.** Prove Theorem 3.

17. Prove this identity relating the Fibonacci numbers and the binomial coefficients:

$$f_{n+1} = C(n, 0) + C(n-1, 1) + \cdots + C(n-k, k),$$

where n is a positive integer and $k = \lfloor n/2 \rfloor$. [*Hint:* Let $a_n = C(n, 0) + C(n-1, 1) + \cdots + C(n-k, k)$. Show that the sequence $\{a_n\}$ satisfies the same recurrence relation and initial conditions satisfied by the sequence of Fibonacci numbers.]

18. Solve the recurrence relation $a_n = 6a_{n-1} - 12a_{n-2} + 8a_{n-3}$ with $a_0 = -5$, $a_1 = 4$, and $a_2 = 88$.

19. Solve the recurrence relation $a_n = -3a_{n-1} - 3a_{n-2} - a_{n-3}$ with $a_0 = 5$, $a_1 = -9$, and $a_2 = 15$.

20. Find the general form of the solutions of the recurrence relation $a_n = 8a_{n-2} - 16a_{n-4}$.

21. What is the general form of the solutions of a linear homogeneous recurrence relation if its characteristic equation has roots 1, 1, 1, 1, −2, −2, −2, 3, 3, −4?

22. What is the general form of the solutions of a linear homogeneous recurrence relation if its characteristic equation has the roots −1, −1, −1, 2, 2, 5, 5, 7?

23. Consider the nonhomogeneous linear recurrence relation $a_n = 3a_{n-1} + 2^n$.

a) Show that $a_n = -2^{n+1}$ is a solution of this recurrence relation.

b) Use Theorem 5 to find all solutions of this recurrence relation.

c) Find the solution with $a_0 = 1$.

24. Consider the nonhomogeneous linear recurrence relation $a_n = 2a_{n-1} + 2^n$.

a) Show that $a_n = n2^n$ is a solution of this recurrence relation.

b) Use Theorem 5 to find all solutions of this recurrence relation.

c) Find the solution with $a_0 = 2$.

25. a) Determine values of the constants A and B such that $a_n = An + B$ is a solution of recurrence relation $a_n = 2a_{n-1} + n + 5$.

b) Use Theorem 5 to find all solutions of this recurrence relation.

c) Find the solution of this recurrence relation with $a_0 = 4$.

26. What is the general form of the particular solution guaranteed to exist by Theorem 6 of the linear nonhomogeneous recurrence relation $a_n = 6a_{n-1} - 12a_{n-2} + 8a_{n-3} + F(n)$ if

a) $F(n) = n^2$? **b)** $F(n) = 2^n$?
c) $F(n) = n2^n$? **d)** $F(n) = (-2)^n$?
e) $F(n) = n^2 2^n$? **f)** $F(n) = n^3(-2)^n$?
g) $F(n) = 3$?

27. What is the general form of the particular solution guaranteed to exist by Theorem 6 of the linear nonhomogeneous recurrence relation $a_n = 8a_{n-2} - 16a_{n-4} + F(n)$ if

a) $F(n) = n^3$? **b)** $F(n) = (-2)^n$?
c) $F(n) = n2^n$? **d)** $F(n) = n^2 4^n$?
e) $F(n) = (n^2 - 2)(-2)^n$? **f)** $F(n) = n^4 2^n$?
g) $F(n) = 2$?

28. a) Find all solutions of the recurrence relation $a_n = 2a_{n-1} + 2n^2$.
b) Find the solution of the recurrence relation in part (a) with initial condition $a_1 = 4$.

29. a) Find all solutions of the recurrence relation $a_n = 2a_{n-1} + 3^n$.
b) Find the solution of the recurrence relation in part (a) with initial condition $a_1 = 5$.

30. a) Find all solutions of the recurrence relation $a_n = -5a_{n-1} - 6a_{n-2} + 42 \cdot 4^n$.
b) Find the solution of this recurrence relation with $a_1 = 56$ and $a_2 = 278$.

31. Find all solutions of the recurrence relation $a_n = 5a_{n-1} - 6a_{n-2} + 2^n + 3n$. [*Hint:* Look for a particular solution of the form $qn2^n + p_1 n + p_2$, where q, p_1, and p_2 are constants.]

32. Find the solution of the recurrence relation $a_n = 2a_{n-1} + 3 \cdot 2^n$.

33. Find all solutions of the recurrence relation $a_n = 4a_{n-1} - 4a_{n-2} + (n+1)2^n$.

34. Find all solutions of the recurrence relation $a_n = 7a_{n-1} - 16a_{n-2} + 12a_{n-3} + n4^n$ with $a_0 = -2$, $a_1 = 0$, and $a_2 = 5$.

35. Find the solution of the recurrence relation $a_n = 4a_{n-1} - 3a_{n-2} + 2^n + n + 3$ with $a_0 = 1$ and $a_1 = 4$.

36. Let a_n be the sum of the first n perfect squares, that is, $a_n = \sum_{k=1}^{n} k^2$. Show that the sequence $\{a_n\}$ satisfies the linear nonhomogeneous recurrence relation $a_n = a_{n-1} + n^2$ and the initial condition $a_1 = 1$. Use Theorem 6 to determine a formula for a_n by solving this recurrence relation.

37. Let a_n be the sum of the first n triangular numbers, that is, $a_n = \sum_{k=1}^{n} t_k$, where $t_k = k(k+1)/2$. Show that $\{a_n\}$ satisfies the linear nonhomogeneous recurrence relation $a_n = a_{n-1} + n(n+1)/2$ and the initial condition $a_1 = 1$. Use Theorem 6 to determine a formula for a_n by solving this recurrence relation.

38. a) Find the characteristic roots of the linear homogeneous recurrence relation $a_n = 2a_{n-1} - 2a_{n-2}$. [*Note:* These are complex numbers.]
b) Find the solution of the recurrence relation in part (a) with $a_0 = 1$ and $a_1 = 2$.

***39. a)** Find the characteristic roots of the linear homogeneous recurrence relation $a_n = a_{n-4}$. [*Note:* These include complex numbers.]
b) Find the solution of the recurrence relation in part (a) with $a_0 = 1$, $a_1 = 0$, $a_2 = -1$, and $a_3 = 1$.

***40.** Solve the simultaneous recurrence relations

$$a_n = 3a_{n-1} + 2b_{n-1}$$
$$b_n = a_{n-1} + 2b_{n-1}$$

with $a_0 = 1$ and $b_0 = 2$.

***41. a)** Use the formula found in Example 4 for f_n, the nth Fibonacci number, to show that f_n is the integer closest to

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

b) Determine for which n f_n is greater than

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$$

and for which n f_n is less than

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

42. Show that if $a_n = a_{n-1} + a_{n-2}$, $a_0 = s$ and $a_1 = t$, where s and t are constants, then $a_n = sf_{n-1} + tf_n$ for all positive integers n.

43. Express the solution of the linear nonhomogenous recurrence relation $a_n = a_{n-1} + a_{n-2} + 1$ for $n \geq 2$ where $a_0 = 0$ and $a_1 = 1$ in terms of the Fibonacci numbers. [*Hint:* Let $b_n = a_n + 1$ and apply Exercise 42 to the sequence b_n.]

***44.** (*Linear algebra required*) Let $\mathbf{A}_n$ be the $n \times n$ matrix with 2s on its main diagonal, 1s in all positions next to a diagonal element, and 0s everywhere else. Find a recurrence relation for d_n, the determinant of $\mathbf{A}_n$. Solve this recurrence relation to find a formula for d_n.

45. Suppose that each pair of a genetically engineered species of rabbits left on an island produces two new pairs of rabbits at the age of 1 month and six new pairs of rabbits at the age of 2 months and every month afterward. None of the rabbits ever die or leave the island.
a) Find a recurrence relation for the number of pairs of rabbits on the island n months after one newborn pair is left on the island.
b) By solving the recurrence relation in (a) determine the number of pairs of rabbits on the island n months after one pair is left on the island.

46. Suppose that there are two goats on an island initially. The number of goats on the island doubles every year by natural reproduction, and some goats are either added or removed each year.

a) Construct a recurrence relation for the number of goats on the island at the start of the nth year, assuming that during each year an extra 100 goats are put on the island.
b) Solve the recurrence relation from part (a) to find the number of goats on the island at the start of the nth year.
c) Construct a recurrence relation for the number of goats on the island at the start of the nth year, assuming that n goats are removed during the nth year for each $n \geq 3$.
d) Solve the recurrence relation in part (c) for the number of goats on the island at the start of the nth year.

47. A new employee at an exciting new software company starts with a salary of \$50,000 and is promised that at the end of each year her salary will be double her salary of the previous year, with an extra increment of \$10,000 for each year she has been with the company.

a) Construct a recurrence relation for her salary for her nth year of employment.
b) Solve this recurrence relation to find her salary for her nth year of employment.

Some linear recurrence relations that do not have constant coefficients can be systematically solved. This is the case for recurrence relations of the form $f(n)a_n = g(n)a_{n-1} + h(n)$. Exercises 48–50 illustrate this.

*__48.__ a) Show that the recurrence relation

$$f(n)a_n = g(n)a_{n-1} + h(n),$$

for $n \geq 1$, and with $a_0 = C$, can be reduced to a recurrence relation of the form

$$b_n = b_{n-1} + Q(n)h(n),$$

where $b_n = g(n+1)Q(n+1)a_n$, with

$$Q(n) = (f(1)f(2)\cdots f(n-1))/(g(1)g(2)\cdots g(n)).$$

b) Use part (a) to solve the original recurrence relation to obtain

$$a_n = \frac{C + \sum_{i=1}^{n} Q(i)h(i)}{g(n+1)Q(n+1)}.$$

*__49.__ Use Exercise 48 to solve the recurrence relation $(n+1)a_n = (n+3)a_{n-1} + n$, for $n \geq 1$, with $a_0 = 1$.

50. It can be shown that C_n, the average number of comparisons made by the quick sort algorithm (described in preamble to Exercise 50 in Section 5.4), when sorting n elements in random order, satisfies the recurrence relation

$$C_n = n + 1 + \frac{2}{n}\sum_{k=0}^{n-1} C_k$$

for $n = 1, 2, \ldots,$ with initial condition $C_0 = 0$.

a) Show that $\{C_n\}$ also satisfies the recurrence relation $nC_n = (n+1)C_{n-1} + 2n$ for $n = 1, 2, \ldots$.
b) Use Exercise 48 to solve the recurrence relation in part (a) to find an explicit formula for C_n.

**__51.__ Prove Theorem 4.

**__52.__ Prove Theorem 6.

53. Solve the recurrence relation $T(n) = nT^2(n/2)$ with initial condition $T(1) = 6$ when $n = 2^k$ for some integer k. [*Hint:* Let $n = 2^k$ and then make the substitution $a_k = \log T(2^k)$ to obtain a linear nonhomogeneous recurrence relation.]

8.3 Divide-and-Conquer Algorithms and Recurrence Relations

8.3.1 Introduction

Links

Many recursive algorithms take a problem with a given input and divide it into one or more smaller problems. This reduction is successively applied until the solutions of the smaller problems can be found quickly. For instance, we perform a binary search by reducing the search for an element in a list to the search for this element in a list half as long. We successively apply this reduction until one element is left. When we sort a list of integers using the merge sort, we split the list into two halves of equal size and sort each half separately. We then merge the two sorted halves. Another example of this type of recursive algorithm is a procedure for multiplying integers that reduces the problem of the multiplication of two integers to three multiplications of pairs of integers with half as many bits. This reduction is successively applied until integers with one bit are obtained. There procedures follow an important algorithmic paradigm known as **divide-and-conquer**, and are called **divide-and-conquer algorithms**, because they *divide* a problem into one or more instances of the same problem of smaller size and they *conquer* the problem by using the solutions of the smaller problems to find a solution of the original problem, perhaps with some additional work.

"Divide et impera" (translation: *"Divide and conquer"*) —Julius Caesar

In this section we will show how recurrence relations can be used to analyze the computational complexity of divide-and-conquer algorithms. We will use these recurrence relations to estimate the number of operations used by many different divide-and-conquer algorithms, including several that we introduce in this section.

8.3.2 Divide-and-Conquer Recurrence Relations

Suppose that a recursive algorithm divides a problem of size n into a subproblems, where each subproblem is of size n/b (for simplicity, assume that n is a multiple of b; in reality, the smaller problems are often of size equal to the nearest integers either less than or equal to, or greater than or equal to, n/b). Also, suppose that a total of $g(n)$ extra operations are required in the conquer step of the algorithm to combine the solutions of the subproblems into a solution of the original problem. Then, if $f(n)$ represents the number of operations required to solve the problem of size n, it follows that f satisfies the recurrence relation

$$f(n) = af(n/b) + g(n).$$

This is called a **divide-and-conquer recurrence relation**.

We will first set up the divide-and-conquer recurrence relations that can be used to study the complexity of some important algorithms. Then we will show how to use these divide-and-conquer recurrence relations to estimate the complexity of these algorithms.

EXAMPLE 1 **Binary Search** We introduced a binary search algorithm in Section 3.1. This binary search algorithm reduces the search for an element in a search sequence of size n to the binary search for this element in a search sequence of size $n/2$, when n is even. (Hence, the problem of size n has been reduced to *one* problem of size $n/2$.) Two comparisons are needed to implement this reduction (one to determine which half of the list to use and the other to determine whether any terms of the list remain). Hence, if $f(n)$ is the number of comparisons required to search for an element in a search sequence of size n, then

Extra Examples

$$f(n) = f(n/2) + 2$$

when n is even. ◀

EXAMPLE 2 **Finding the Maximum and Minimum of a Sequence** Consider the following algorithm for locating the maximum and minimum elements of a sequence $a_1, a_2, \ldots, a_n$. If $n = 1$, then a_1 is the maximum and the minimum. If $n > 1$, split the sequence into two sequences, either where both have the same number of elements or where one of the sequences has one more element than the other. The problem is reduced to finding the maximum and minimum of each of the two smaller sequences. The solution to the original problem results from the comparison of the separate maxima and minima of the two smaller sequences to obtain the overall maximum and minimum.

Let $f(n)$ be the total number of comparisons needed to find the maximum and minimum elements of the sequence with n elements. We have shown that a problem of size n can be reduced into two problems of size $n/2$, when n is even, using two comparisons, one to compare the maxima of the two sequences and the other to compare the minima of the two sequences. This gives the recurrence relation

$$f(n) = 2f(n/2) + 2$$

when n is even. ◀

EXAMPLE 3 **Merge Sort** The merge sort algorithm (introduced in Section 5.4) splits a list to be sorted with n items, where n is even, into two lists with $n/2$ elements each, and uses fewer than n

comparisons to merge the two sorted lists of $n/2$ items each into one sorted list. Consequently, the number of comparisons used by the merge sort to sort a list of n elements is less than $M(n)$, where the function $M(n)$ satisfies the divide-and-conquer recurrence relation

$$M(n) = 2M(n/2) + n.$$

◀

EXAMPLE 4 **Fast Multiplication of Integers** Surprisingly, there are more efficient algorithms than the conventional algorithm (described in Section 4.2) for multiplying integers. One of these algorithms, which uses a divide-and-conquer technique, will be described here. This fast multiplication algorithm proceeds by splitting each of two $2n$-bit integers into two blocks, each with n bits. Then, the original multiplication is reduced from the multiplication of two $2n$-bit integers to three multiplications of n-bit integers, plus shifts and additions.

Links

Suppose that a and b are integers with binary expansions of length $2n$ (add initial bits of zero in these expansions if necessary to make them the same length). Let

$$a = (a_{2n-1}a_{2n-2}\cdots a_1a_0)_2 \qquad \text{and} \qquad b = (b_{2n-1}b_{2n-2}\cdots b_1b_0)_2.$$

Let

$$a = 2^nA_1 + A_0, \quad b = 2^nB_1 + B_0,$$

where

$$A_1 = (a_{2n-1}\cdots a_{n+1}a_n)_2, \qquad A_0 = (a_{n-1}\cdots a_1a_0)_2,$$
$$B_1 = (b_{2n-1}\cdots b_{n+1}b_n)_2, \qquad B_0 = (b_{n-1}\cdots b_1b_0)_2.$$

The algorithm for fast multiplication of integers is based on the fact that ab can be rewritten as

$$ab = (2^{2n} + 2^n)A_1B_1 + 2^n(A_1 - A_0)(B_0 - B_1) + (2^n + 1)A_0B_0.$$

The important fact about this identity is that it shows that the multiplication of two $2n$-bit integers can be carried out using three multiplications of n-bit integers, together with additions, subtractions, and shifts. This shows that if $f(n)$ is the total number of bit operations needed to multiply two n-bit integers, then

$$f(2n) = 3f(n) + Cn.$$

The reasoning behind this equation is as follows. The three multiplications of n-bit integers are carried out using $3f(n)$-bit operations. Each of the additions, subtractions, and shifts uses a constant multiple of n-bit operations, and Cn represents the total number of bit operations used by these operations. ◀

EXAMPLE 5 **Fast Matrix Multiplication** In Example 7 of Section 3.3 we showed that multiplying two $n \times n$ matrices using the definition of matrix multiplication required n^3 multiplications and $n^2(n-1)$ additions. Consequently, computing the product of two $n \times n$ matrices in this way requires $O(n^3)$ operations (multiplications and additions). Surprisingly, there are more efficient divide-and-conquer algorithms for multiplying two $n \times n$ matrices. Such an algorithm, invented by Volker Strassen in 1969, reduces the multiplication of two $n \times n$ matrices, when n is even, to seven multiplications of two $(n/2) \times (n/2)$ matrices and 15 additions of $(n/2) \times (n/2)$ matrices.

Links

(See [CoLeRiSt09] for the details of this algorithm.) Hence, if $f(n)$ is the number of operations (multiplications and additions) used, it follows that

$$f(n) = 7f(n/2) + 15n^2/4$$

when n is even. ◀

As Examples 1–5 show, recurrence relations of the form $f(n) = af(n/b) + g(n)$ arise in many different situations. It is possible to derive estimates of the size of functions that satisfy such recurrence relations. Suppose that f satisfies this recurrence relation whenever n is divisible by b. Let $n = b^k$, where k is a positive integer. Then

$$\begin{aligned} f(n) &= af(n/b) + g(n) \\ &= a^2 f(n/b^2) + ag(n/b) + g(n) \\ &= a^3 f(n/b^3) + a^2 g(n/b^2) + ag(n/b) + g(n) \\ &\ \vdots \\ &= a^k f(n/b^k) + \sum_{j=0}^{k-1} a^j g(n/b^j). \end{aligned}$$

Because $n/b^k = 1$, it follows that

$$f(n) = a^k f(1) + \sum_{j=0}^{k-1} a^j g(n/b^j).$$

We can use this equation for $f(n)$ to estimate the size of functions that satisfy divide-and-conquer relations.

THEOREM 1 Let f be an increasing function that satisfies the recurrence relation

$$f(n) = af(n/b) + c$$

whenever n is divisible by b, where $a \geq 1$, b is an integer greater than 1, and c is a positive real number. Then

$$f(n) \text{ is } \begin{cases} O(n^{\log_b a}) & \text{if } a > 1, \\ O(\log n) & \text{if } a = 1. \end{cases}$$

Furthermore, when $n = b^k$ and $a \neq 1$, where k is a positive integer,

$$f(n) = C_1 n^{\log_b a} + C_2,$$

where $C_1 = f(1) + c/(a-1)$ and $C_2 = -c/(a-1)$.

Proof: First let $n = b^k$. From the expression for $f(n)$ obtained in the discussion preceding the theorem, with $g(n) = c$, we have

$$f(n) = a^k f(1) + \sum_{j=0}^{k-1} a^j c = a^k f(1) + c \sum_{j=0}^{k-1} a^j.$$

When $a = 1$ we have

$$f(n) = f(1) + ck.$$

Because $n = b^k$, we have $k = \log_b n$. Hence,

$$f(n) = f(1) + c\log_b n.$$

When n is not a power of b, we have $b^k < n < b^{k+1}$, for a positive integer k. Because f is increasing, it follows that $f(n) \le f(b^{k+1}) = f(1) + c(k+1) = (f(1)+c) + ck \le (f(1)+c) + c\log_b n$. Therefore, in both cases, $f(n)$ is $O(\log n)$ when $a = 1$.

Now suppose that $a > 1$. First assume that $n = b^k$, where k is a positive integer. From the formula for the sum of terms of a geometric progression (Theorem 1 in Section 2.4), it follows that

$$\begin{aligned} f(n) &= a^k f(1) + c(a^k - 1)/(a-1) \\ &= a^k[f(1) + c/(a-1)] - c/(a-1) \\ &= C_1 n^{\log_b a} + C_2, \end{aligned}$$

because $a^k = a^{\log_b n} = n^{\log_b a}$ (see Exercise 4 in Appendix 2), where $C_1 = f(1) + c/(a-1)$ and $C_2 = -c/(a-1)$.

Now suppose that n is not a power of b. Then $b^k < n < b^{k+1}$, where k is a nonnegative integer. Because f is increasing,

$$\begin{aligned} f(n) &\le f(b^{k+1}) = C_1 a^{k+1} + C_2 \\ &\le (C_1 a) a^{\log_b n} + C_2 \\ &= (C_1 a) n^{\log_b a} + C_2, \end{aligned}$$

because $k \le \log_b n < k+1$.

Hence, we have $f(n)$ is $O(n^{\log_b a})$. ◁

Examples 6–9 illustrate how Theorem 1 is used.

EXAMPLE 6 *Extra Examples* Let $f(n) = 5f(n/2) + 3$ and $f(1) = 7$. Find $f(2^k)$, where k is a positive integer. Also, estimate $f(n)$ if f is an increasing function.

Solution: From the proof of Theorem 1, with $a = 5$, $b = 2$, and $c = 3$, we see that if $n = 2^k$, then

$$\begin{aligned} f(n) &= a^k[f(1) + c/(a-1)] + [-c/(a-1)] \\ &= 5^k[7 + (3/4)] - 3/4 \\ &= 5^k(31/4) - 3/4. \end{aligned}$$

Also, if $f(n)$ is increasing, Theorem 1 shows that $f(n)$ is $O(n^{\log_b a}) = O(n^{\log 5})$. ◀

We can use Theorem 1 to estimate the computational complexity of the binary search algorithm and the algorithm given in Example 2 for locating the minimum and maximum of a sequence.

EXAMPLE 7 Give a big-O estimate for the number of comparisons used by a binary search.

Solution: In Example 1 it was shown that $f(n) = f(n/2) + 2$ when n is even, where f is the number of comparisons required to perform a binary search on a sequence of size n. Hence, from Theorem 1, it follows that $f(n)$ is $O(\log n)$. ◀

EXAMPLE 8 Give a big-O estimate for the number of comparisons used to locate the maximum and minimum elements in a sequence using the algorithm given in Example 2.

Solution: In Example 2 we showed that $f(n) = 2f(n/2) + 2$, when n is even, where f is the number of comparisons needed by this algorithm. Hence, from Theorem 1, it follows that $f(n)$ is $O(n^{\log 2}) = O(n)$. ◀

We now state a more general, and more complicated, theorem, which has Theorem 1 as a special case. This theorem (or more powerful versions, including big-Theta estimates) is sometimes known as the master theorem because it is useful in analyzing the complexity of many important divide-and-conquer algorithms.

THEOREM 2 **MASTER THEOREM** Let f be an increasing function that satisfies the recurrence relation

$$f(n) = af(n/b) + cn^d$$

whenever $n = b^k$, where k is a positive integer, $a \geq 1$, b is an integer greater than 1, and c and d are real numbers with c positive and d nonnegative. Then

$$f(n) \text{ is } \begin{cases} O(n^d) & \text{if } a < b^d, \\ O(n^d \log n) & \text{if } a = b^d, \\ O(n^{\log_b a}) & \text{if } a > b^d. \end{cases}$$

The proof of Theorem 2 is left for the reader as Exercises 29–33.

EXAMPLE 9 **Complexity of Merge Sort** In Example 3 we explained that the number of comparisons used by the merge sort to sort a list of n elements is less than $M(n)$, where $M(n) = 2M(n/2) + n$. By the master theorem (Theorem 2) we find that $M(n)$ is $O(n \log n)$, which agrees with the estimate found in Section 5.4. ◀

EXAMPLE 10 Give a big-O estimate for the number of bit operations needed to multiply two n-bit integers using the fast multiplication algorithm described in Example 4.

Solution: Example 4 shows that $f(n) = 3f(n/2) + Cn$, when n is even, where $f(n)$ is the number of bit operations required to multiply two n-bit integers using the fast multiplication algorithm. Hence, from the master theorem (Theorem 2), it follows that $f(n)$ is $O(n^{\log 3})$. Note that log 3 $\sim$ 1.6. Because the conventional algorithm for multiplication uses $O(n^2)$ bit operations, the fast multiplication algorithm is a substantial improvement over the conventional algorithm in terms of time complexity for sufficiently large integers, including large integers that occur in practical applications. ◀

EXAMPLE 11 Give a big-O estimate for the number of multiplications and additions required to multiply two $n \times n$ matrices using the matrix multiplication algorithm referred to in Example 5.

Solution: Let $f(n)$ denote the number of additions and multiplications used by the algorithm mentioned in Example 5 to multiply two $n \times n$ matrices. We have $f(n) = 7f(n/2) + 15n^2/4$, when n is even. Hence, from the master theorem (Theorem 2), it follows that $f(n)$ is $O(n^{\log 7})$. Note that $\log 7 \sim 2.8$. Because the conventional algorithm for multiplying two $n \times n$ matrices uses $O(n^3)$ additions and multiplications, it follows that for sufficiently large integers n, including those that occur in many practical applications, this algorithm is substantially more efficient in time complexity than the conventional algorithm. ◀

THE CLOSEST-PAIR PROBLEM We conclude this section by introducing a divide-and-conquer algorithm from computational geometry, the part of discrete mathematics devoted to algorithms that solve geometric problems.

EXAMPLE 12 **The Closest-Pair Problem** Consider the problem of determining the closest pair of points in a set of n points $(x_1, y_1), \ldots, (x_n, y_n)$ in the plane, where the distance between two points (x_i, y_i) and (x_j, y_j) is the usual Euclidean distance $\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. This problem arises in many applications such as determining the closest pair of airplanes in the air space at a particular altitude being managed by an air traffic controller. How can this closest pair of points be found in an efficient way?

Links

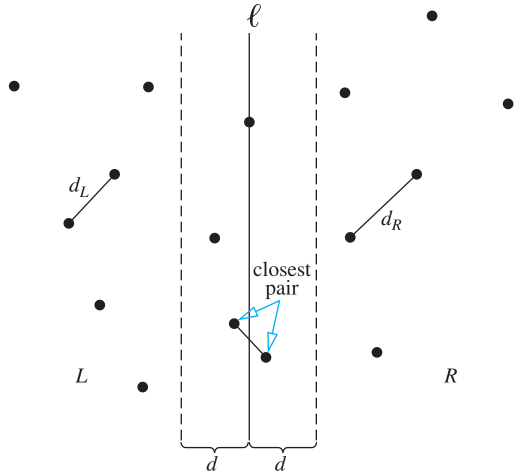
Solution: To solve this problem we can first determine the distance between every pair of points and then find the smallest of these distances. However, this approach requires $O(n^2)$ computations of distances and comparisons because there are $C(n, 2) = n(n-1)/2$ pairs of points. Surprisingly, there is an elegant divide-and-conquer algorithm that can solve the closest-pair problem for n points using $O(n \log n)$ computations of distances and comparisons. The algorithm we describe here is due to Michael Samos (see [PrSa85]).

It took researchers more than 10 years to find an algorithm with $O(n \log n)$ complexity that locates the closest pair of points among n points.

For simplicity, we assume that $n = 2^k$, where k is a positive integer. (We avoid some technical considerations that are needed when n is not a power of 2.) When $n = 2$, we have only one pair of points; the distance between these two points is the minimum distance. At the start of the algorithm we use the merge sort twice, once to sort the points in order of increasing x coordinates, and once to sort the points in order of increasing y coordinates. Each of these sorts requires $O(n \log n)$ operations. We will use these sorted lists in each recursive step.

The recursive part of the algorithm divides the problem into two subproblems, each involving half as many points. Using the sorted list of the points by their x coordinates, we construct a vertical line ℓ dividing the n points into two parts, a left part and a right part of equal size, each containing $n/2$ points, as shown in Figure 1. (If any points fall on the dividing line ℓ, we divide them among the two parts if necessary.) At subsequent steps of the recursion we need not sort on x coordinates again, because we can select the corresponding sorted subset of all the points. This selection is a task that can be done with $O(n)$ comparisons.

There are three possibilities concerning the positions of the closest points: (1) they are both in the left region L, (2) they are both in the right region R, or (3) one point is in the left region and the other is in the right region. Apply the algorithm recursively to compute d_L and d_R, where d_L is the minimum distance between points in the left region and d_R is the minimum distance between points in the right region. Let $d = \min(d_L, d_R)$. To successfully divide the problem of finding the closest two points in the original set into the two problems of finding the



In this illustration the problem of finding the closest pair in a set of 16 points is reduced to two problems of finding the closest pair in a set of eight points *and* the problem of determining whether there are points closer than $d = \min(d_L, d_R)$ within the strip of width $2d$ centered at ℓ.

FIGURE 1 The recursive step of the algorithm for solving the closest-pair problem.

shortest distances between points in the two regions separately, we have to handle the conquer part of the algorithm, which requires that we consider the case where the closest points lie in different regions, that is, one point is in L and the other in R. Because there is a pair of points at distance d where both points lie in R or both points lie in L, for the closest points to lie in different regions requires that they must be a distance less than d apart.

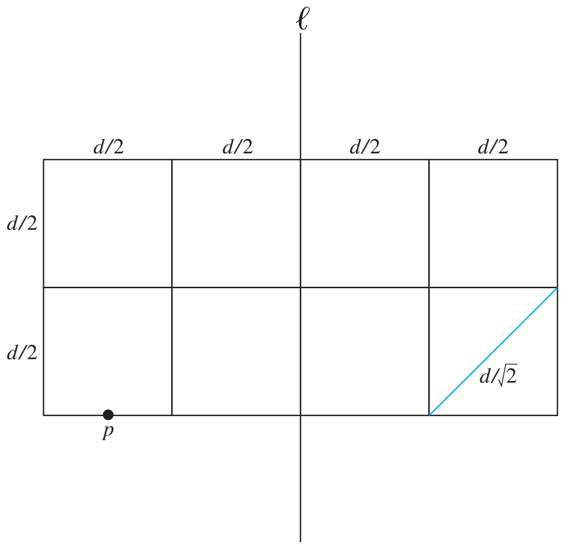
For a point in the left region and a point in the right region to lie at a distance less than d apart, these points must lie in the vertical strip of width $2d$ that has the line ℓ as its center. (Otherwise, the distance between these points is greater than the difference in their x coordinates, which exceeds d.) To examine the points within this strip, we sort the points so that they are listed in order of increasing y coordinates, using the sorted list of the points by their y coordinates. At each recursive step, we form a subset of the points in the region sorted by their y coordinates from the already sorted set of all points sorted by their y coordinates, which can be done with $O(n)$ comparisons.

Beginning with a point in the strip with the smallest y coordinate, we successively examine each point in the strip, computing the distance between this point and all other points in the strip that have larger y coordinates that could lie at a distance less than d from this point. Note that to examine a point p, we need only consider the distances between p and points in the set that lie within the rectangle of height d and width $2d$ with p on its base and with vertical sides at distance d from ℓ.

We can show that there are at most eight points from the set, including p, in or on this $2d \times d$ rectangle. To see this, note that there can be at most one point in each of the eight $d/2 \times d/2$ squares shown in Figure 2. This follows because the farthest apart points can be on or within one of these squares is the diagonal length $d/\sqrt{2}$ (which can be found using the Pythagorean theorem), which is less than d, and each of these $d/2 \times d/2$ squares lies entirely within the left region or the right region. This means that at this stage we need only compare at most seven distances, the distances between p and the seven or fewer other points in or on the rectangle, with d.

Because the total number of points in the strip of width $2d$ does not exceed n (the total number of points in the set), at most $7n$ distances need to be compared with d to find the minimum distance between points. That is, there are only $7n$ possible distances that could be less than d. Consequently, once the merge sort has been used to sort the pairs according to their x coordinates and according to their y coordinates, we find that the increasing function $f(n)$ satisfying the recurrence relation

$$f(n) = 2f(n/2) + 7n,$$



At most eight points, including p, can lie in or on the $2d \times d$ rectangle centered at ℓ because at most one point can lie in or on each of the eight $(d/2) \times (d/2)$ squares.

FIGURE 2 Showing that there are at most seven other points to consider for each point in the strip.

where $f(2) = 1$, exceeds the number of comparisons needed to solve the closest-pair problem for n points. By the master theorem (Theorem 2), it follows that $f(n)$ is $O(n \log n)$. The two sorts of points by their x coordinates and by their y coordinates each can be done using $O(n \log n)$ comparisons, by using the merge sort, and the sorted subsets of these coordinates at each of the $O(\log n)$ steps of the algorithm can be done using $O(n)$ comparisons each. Thus, we find that the closest-pair problem can be solved using $O(n \log n)$ comparisons. ◀

Exercises

1. How many comparisons are needed for a binary search in a set of 64 elements?

2. How many comparisons are needed to locate the maximum and minimum elements in a sequence with 128 elements using the algorithm in Example 2?

3. Multiply $(1110)_2$ and $(1010)_2$ using the fast multiplication algorithm.

4. Express the fast multiplication algorithm in pseudocode.

5. Determine a value for the constant C in Example 4 and use it to estimate the number of bit operations needed to multiply two 64-bit integers using the fast multiplication algorithm.

6. How many operations are needed to multiply two 32×32 matrices using the algorithm referred to in Example 5?

7. Suppose that $f(n) = f(n/3) + 1$ when n is a positive integer divisible by 3, and $f(1) = 1$. Find

a) $f(3)$. **b)** $f(27)$. **c)** $f(729)$.

8. Suppose that $f(n) = 2f(n/2) + 3$ when n is an even positive integer, and $f(1) = 5$. Find

a) $f(2)$. **b)** $f(8)$. **c)** $f(64)$. **d)** $f(1024)$.

9. Suppose that $f(n) = f(n/5) + 3n^2$ when n is a positive integer divisible by 5, and $f(1) = 4$. Find

a) $f(5)$. **b)** $f(125)$. **c)** $f(3125)$.

10. Find $f(n)$ when $n = 2^k$, where f satisfies the recurrence relation $f(n) = f(n/2) + 1$ with $f(1) = 1$.

11. Give a big-O estimate for the function f in Exercise 10 if f is an increasing function.

12. Find $f(n)$ when $n = 3^k$, where f satisfies the recurrence relation $f(n) = 2f(n/3) + 4$ with $f(1) = 1$.

13. Give a big-O estimate for the function f in Exercise 12 if f is an increasing function.

14. Suppose that there are $n = 2^k$ teams in an elimination tournament, where there are $n/2$ games in the first round, with the $n/2 = 2^{k-1}$ winners playing in the second round, and so on. Develop a recurrence relation for the number of rounds in the tournament.

15. How many rounds are in the elimination tournament described in Exercise 14 when there are 32 teams?

16. Solve the recurrence relation for the number of rounds in the tournament described in Exercise 14.

17. Suppose that the votes of n people for different candidates (where there can be more than two candidates) for a particular office are the elements of a sequence. A person wins the election if this person receives a majority of the votes.

a) Devise a divide-and-conquer algorithm that determines whether a candidate received a majority and, if so, determine who this candidate is. [*Hint:* Assume that n is even and split the sequence of votes into two sequences, each with $n/2$ elements. Note that a candidate could not have received a majority of votes without receiving a majority of votes in at least one of the two halves.]

b) Use the master theorem to give a big-O estimate for the number of comparisons needed by the algorithm you devised in part (a).

18. Suppose that each person in a group of n people votes for exactly two people from a slate of candidates to fill two positions on a committee. The top two finishers both win positions as long as each receives more than $n/2$ votes.

a) Devise a divide-and-conquer algorithm that determines whether the two candidates who received the most votes each received at least $n/2$ votes and, if so, determine who these two candidates are.

b) Use the master theorem to give a big-O estimate for the number of comparisons needed by the algorithm you devised in part (a).

19. a) Set up a divide-and-conquer recurrence relation for the number of multiplications required to compute x^n, where x is a real number and n is a positive integer, using the recursive algorithm from Exercise 26 in Section 5.4.

b) Use the recurrence relation you found in part (a) to construct a big-O estimate for the number of multiplications used to compute x^n using the recursive algorithm.

20. a) Set up a divide-and-conquer recurrence relation for the number of modular multiplications required to compute a^n **mod** m, where a, m, and n are positive integers, using the recursive algorithms from Example 4 in Section 5.4.

b) Use the recurrence relation you found in part (a) to construct a big-O estimate for the number of modular multiplications used to compute a^n **mod** m using the recursive algorithm.

21. Suppose that the function f satisfies the recurrence relation $f(n) = 2f(\sqrt{n}) + 1$ whenever n is a perfect square greater than 1 and $f(2) = 1$.

a) Find $f(16)$.

b) Give a big-O estimate for $f(n)$. [*Hint:* Make the substitution $m = \log n$.]

22. Suppose that the function f satisfies the recurrence relation $f(n) = 2f(\sqrt{n}) + \log n$ whenever n is a perfect square greater than 1 and $f(2) = 1$.

a) Find $f(16)$.

b) Find a big-O estimate for $f(n)$. [*Hint:* Make the substitution $m = \log n$.]

****23.** This exercise deals with the problem of finding the largest sum of consecutive terms of a sequence of n real numbers. When all terms are positive, the sum of all terms provides the answer, but the situation is more complicated when some terms are negative. For example, the maximum sum of consecutive terms of the sequence $-2, 3, -1, 6, -7, 4$ is $3 + (-1) + 6 = 8$. (This exercise is based on [Be86].) Recall that in Exercise 56 in Section 8.1 we developed a dynamic programming algorithm for solving this problem. Here, we first look at the brute-force algorithm for solving this problem; then we develop a divide-and-conquer algorithm for solving it.

a) Use pseudocode to describe an algorithm that solves this problem by finding the sums of consecutive terms starting with the first term, the sums of consecutive terms starting with the second term, and so on, keeping track of the maximum sum found so far as the algorithm proceeds.

b) Determine the computational complexity of the algorithm in part (a) in terms of the number of sums computed and the number of comparisons made.

c) Devise a divide-and-conquer algorithm to solve this problem. [*Hint:* Assume that there are an even number of terms in the sequence and split the sequence into two halves. Explain how to handle the case when the maximum sum of consecutive terms includes terms in both halves.]

d) Use the algorithm from part (c) to find the maximum sum of consecutive terms of each of the sequences: $-2, 4, -1, 3, 5, -6, 1, 2$; $4, 1, -3, 7, -1, -5, 3, -2$; and $-1, 6, 3, -4, -5, 8, -1, 7$.

e) Find a recurrence relation for the number of sums and comparisons used by the divide-and-conquer algorithm from part (c).

f) Use the master theorem to estimate the computational complexity of the divide-and-conquer algorithm. How does it compare in terms of computational complexity with the algorithm from part (a)?

24. Apply the algorithm described in Example 12 for finding the closest pair of points, using the Euclidean distance between points, to find the closest pair of the points $(1, 3)$, $(1, 7)$, $(2, 4)$, $(2, 9)$, $(3, 1)$, $(3, 5)$, $(4, 3)$, and $(4, 7)$.

25. Apply the algorithm described in Example 12 for finding the closest pair of points, using the Euclidean distance between points, to find the closest pair of the points $(1, 2)$, $(1, 6)$, $(2, 4)$, $(2, 8)$, $(3, 1)$, $(3, 6)$, $(3, 10)$, $(4, 3)$, $(5, 1)$, $(5, 5)$, $(5, 9)$, $(6, 7)$, $(7, 1)$, $(7, 4)$, $(7, 9)$, and $(8, 6)$.

***26.** Use pseudocode to describe the recursive algorithm for solving the closest-pair problem as described in Example 12.

27. Construct a variation of the algorithm described in Example 12 along with justifications of the steps used by the algorithm to find the smallest distance between two points if the distance between two points is defined to be $d((x_i, y_i), (x_j, y_j)) = \max(|x_i - x_j|, |y_i - y_j|)$.

***28.** Suppose someone picks a number x from a set of n numbers. A second person tries to guess the number by successively selecting subsets of the n numbers and asking the first person whether x is in each set. The first person answers either "yes" or "no." When the first

Links person answers each query truthfully, we can find x using $\log n$ queries by successively splitting the sets used in each query in half. Ulam's problem, proposed by Stanislaw Ulam in 1976, asks for the number of queries required to find x, supposing that the first person is allowed to lie exactly once.

a) Show that by asking each question twice, given a number x and a set with n elements, and asking one more question when we find the lie, Ulam's problem can be solved using $2 \log n + 1$ queries.

b) Show that by dividing the initial set of n elements into four parts, each with $n/4$ elements, 1/4 of the elements can be eliminated using two queries. [*Hint:* Use two queries, where each of the queries asks whether the element is in the union of two of the subsets with $n/4$ elements and where one of the subsets of $n/4$ elements is used in both queries.]

c) Show from part (b) that if $f(n)$ equals the number of queries used to solve Ulam's problem using the method from part (b) and n is divisible by 4, then $f(n) = f(3n/4) + 2$.

d) Solve the recurrence relation in part (c) for $f(n)$.

e) Is the naive way to solve Ulam's problem by asking each question twice or the divide-and-conquer method based on part (b) more efficient? The most efficient way to solve Ulam's problem has been determined by A. Pelc [Pe87].

In Exercises 29–33, assume that f is an increasing function satisfying the recurrence relation $f(n) = af(n/b) + cn^d$, where $a \geq 1$, b is an integer greater than 1, and c and d are positive real numbers. These exercises supply a proof of Theorem 2.

***29.** Show that if $a = b^d$ and n is a power of b, then $f(n) = f(1)n^d + cn^d \log_b n$.

30. Use Exercise 29 to show that if $a = b^d$, then $f(n)$ is $O(n^d \log n)$.

***31.** Show that if $a \neq b^d$ and n is a power of b, then $f(n) = C_1 n^d + C_2 n^{\log_b a}$, where $C_1 = b^d c/(b^d - a)$ and $C_2 = f(1) + b^d c/(a - b^d)$.

32. Use Exercise 31 to show that if $a < b^d$, then $f(n)$ is $O(n^d)$.

33. Use Exercise 31 to show that if $a > b^d$, then $f(n)$ is $O(n^{\log_b a})$.

34. Find $f(n)$ when $n = 4^k$, where f satisfies the recurrence relation $f(n) = 5f(n/4) + 6n$, with $f(1) = 1$.

35. Give a big-O estimate for the function f in Exercise 34 if f is an increasing function.

36. Find $f(n)$ when $n = 2^k$, where f satisfies the recurrence relation $f(n) = 8f(n/2) + n^2$ with $f(1) = 1$.

37. Give a big-O estimate for the function f in Exercise 36 if f is an increasing function.

8.4 Generating Functions

8.4.1 Introduction

Links Generating functions are used to represent sequences efficiently by coding the terms of a sequence as coefficients of powers of a variable x in a formal power series. Generating functions can be used to solve many types of counting problems, such as the number of ways to select or distribute objects of different kinds, subject to a variety of constraints, and the number of ways to make change for a dollar using coins of different denominations. Generating functions can be used to solve recurrence relations by translating a recurrence relation for the terms of a sequence into an equation involving a generating function. This equation can then be solved to find a closed form for the generating function. From this closed form, the coefficients of the power series for the generating function can be found, solving the original recurrence relation. Generating functions can also be used to prove combinatorial identities by taking advantage of relatively simple relationships between functions that can be translated into identities involving the terms of sequences. Generating functions are a helpful tool for studying many properties of sequences besides those described in this section, such as their use for establishing asymptotic formulae for the terms of a sequence.

We begin with the definition of the generating function for a sequence.

Definition 1 The *generating function for the sequence* $a_0, a_1, \ldots, a_k, \ldots$ of real numbers is the infinite series

$$G(x) = a_0 + a_1 x + \cdots + a_k x^k + \cdots = \sum_{k=0}^{\infty} a_k x^k.$$

Remark: The generating function for $\{a_k\}$ given in Definition 1 is sometimes called the **ordinary generating function** of $\{a_k\}$ to distinguish it from other types of generating functions for this sequence.

EXAMPLE 1 The generating functions for the sequences $\{a_k\}$ with $a_k = 3$, $a_k = k + 1$, and $a_k = 2^k$ are $\sum_{k=0}^{\infty} 3x^k$, $\sum_{k=0}^{\infty}(k + 1)x^k$, and $\sum_{k=0}^{\infty} 2^k x^k$, respectively. ◀

Extra Examples

We can define generating functions for finite sequences of real numbers by extending a finite sequence $a_0, a_1, \dots, a_n$ into an infinite sequence by setting $a_{n+1} = 0$, $a_{n+2} = 0$, and so on. The generating function $G(x)$ of this infinite sequence $\{a_n\}$ is a polynomial of degree n because no terms of the form $a_j x^j$ with $j > n$ occur, that is,

$$G(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

EXAMPLE 2 What is the generating function for the sequence 1, 1, 1, 1, 1, 1?

Solution: The generating function of 1, 1, 1, 1, 1, 1 is

$$1 + x + x^2 + x^3 + x^4 + x^5.$$

By Theorem 1 of Section 2.4 we have

$$(x^6 - 1)/(x - 1) = 1 + x + x^2 + x^3 + x^4 + x^5$$

when $x \neq 1$. Consequently, $G(x) = (x^6 - 1)/(x - 1)$ is the generating function of the sequence 1, 1, 1, 1, 1, 1. [Because the powers of x are only place holders for the terms of the sequence in a generating function, we do not need to worry that $G(1)$ is undefined.] ◀

EXAMPLE 3 Let m be a positive integer. Let $a_k = C(m, k)$, for $k = 0, 1, 2, \dots, m$. What is the generating function for the sequence $a_0, a_1, \dots, a_m$?

Solution: The generating function for this sequence is

$$G(x) = C(m, 0) + C(m, 1)x + C(m, 2)x^2 + \cdots + C(m, m)x^m.$$

The binomial theorem shows that $G(x) = (1 + x)^m$. ◀

8.4.2 Useful Facts About Power Series

When generating functions are used to solve counting problems, they are usually considered to be **formal power series**. As such, they are treated as algebraic objects; questions about their convergence are ignored. However, when formal power series are convergent, valid operations carry over to their use as formal power series. We will take advantage of the power series of particular functions around $x = 0$. These power series are unique and have a positive radius of convergence. Readers familiar with calculus can consult textbooks on this subject for details about power series, including the convergence of the series we use here.

We will now state some widely important facts about infinite series used when working with generating functions. These facts can be found in calculus texts.

EXAMPLE 4 The function $f(x) = 1/(1-x)$ is the generating function of the sequence $1, 1, 1, 1, \dots$, because

$$1/(1-x) = 1 + x + x^2 + \cdots$$

for $|x| < 1$. ◀

EXAMPLE 5 The function $f(x) = 1/(1-ax)$ is the generating function of the sequence $1, a, a^2, a^3, \dots$, because

$$1/(1-ax) = 1 + ax + a^2x^2 + \cdots$$

when $|ax| < 1$, or equivalently, for $|x| < 1/|a|$ for $a \neq 0$. ◀

We also will need some results on how to add and how to multiply two generating functions. Proofs of these results can be found in calculus texts.

THEOREM 1 Let $f(x) = \sum_{k=0}^{\infty} a_k x^k$ and $g(x) = \sum_{k=0}^{\infty} b_k x^k$. Then

$$f(x) + g(x) = \sum_{k=0}^{\infty} (a_k + b_k)x^k \quad \text{and} \quad f(x)g(x) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} a_j b_{k-j} \right) x^k.$$

Remark: Theorem 1 is valid only for power series that converge in an interval, as all series considered in this section do. However, the theory of generating functions is not limited to such series. In the case of series that do not converge, the statements in Theorem 1 can be taken as definitions of addition and multiplication of generating functions.

We will illustrate how Theorem 1 can be used with Example 6.

EXAMPLE 6 Let $f(x) = 1/(1-x)^2$. Use Example 4 to find the coefficients $a_0, a_1, a_2, \dots$ in the expansion $f(x) = \sum_{k=0}^{\infty} a_k x^k$.

Solution: From Example 4 we see that

$$1/(1-x) = 1 + x + x^2 + x^3 + \cdots.$$

Hence, from Theorem 1, we have

$$1/(1-x)^2 = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} 1 \right) x^k = \sum_{k=0}^{\infty} (k+1)x^k.$$

◀

Remark: This result also can be derived from Example 4 by differentiation. Taking derivatives is a useful technique for producing new identities from existing identities for generating functions.

To use generating functions to solve many important counting problems, we will need to apply the binomial theorem for exponents that are not positive integers. Before we state an extended version of the binomial theorem, we need to define extended binomial coefficients.

Definition 2 Let u be a real number and k a nonnegative integer. Then the *extended binomial coefficient* $\binom{u}{k}$ is defined by

$$\binom{u}{k} = \begin{cases} u(u-1)\cdots(u-k+1)/k! & \text{if } k > 0, \\ 1 & \text{if } k = 0. \end{cases}$$

EXAMPLE 7 Find the values of the extended binomial coefficients $\binom{-2}{3}$ and $\binom{1/2}{3}$.

Solution: Taking $u = -2$ and $k = 3$ in Definition 2 gives us

$$\binom{-2}{3} = \frac{(-2)(-3)(-4)}{3!} = -4.$$

Similarly, taking $u = 1/2$ and $k = 3$ gives us

$$\begin{aligned}\binom{1/2}{3} &= \frac{(1/2)(1/2-1)(1/2-2)}{3!} \\ &= (1/2)(-1/2)(-3/2)/6 \\ &= 1/16.\end{aligned}$$

◀

Example 8 provides a useful formula for extended binomial coefficients when the top parameter is a negative integer. It will be useful in our subsequent discussions.

EXAMPLE 8 When the top parameter is a negative integer, the extended binomial coefficient can be expressed in terms of an ordinary binomial coefficient. To see that this is the case, note that

$$\begin{aligned}\binom{-n}{r} &= \frac{(-n)(-n-1)\cdots(-n-r+1)}{r!} && \text{by definition of extended binomial coefficient} \\ &= \frac{(-1)^r n(n+1)\cdots(n+r-1)}{r!} && \text{factoring out } -1 \text{ from each term in the numerator} \\ &= \frac{(-1)^r (n+r-1)(n+r-2)\cdots n}{r!} && \text{by the commutative law for multiplication} \\ &= \frac{(-1)^r (n+r-1)!}{r!(n-1)!} && \text{multiplying both the numerator and denominator by } (n-1)! \\ &= (-1)^r \binom{n+r-1}{r} && \text{by the definition of binomial coefficients} \\ &= (-1)^r C(n+r-1, r) && \text{using alternative notation for binomial coefficients.}\end{aligned}$$

◀

We now state the extended binomial theorem.

THEOREM 2 **THE EXTENDED BINOMIAL THEOREM** Let x be a real number with $|x| < 1$ and let u be a real number. Then

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

Theorem 2 can be proved using the theory of Maclaurin series. We leave its proof to the reader with a familiarity with this part of calculus.

Remark: When u is a positive integer, the extended binomial theorem reduces to the binomial theorem presented in Section 6.4, because in that case $\binom{u}{k} = 0$ if $k > u$.

Example 9 illustrates the use of Theorem 2 when the exponent is a negative integer.

EXAMPLE 9 Find the generating functions for $(1 + x)^{-n}$ and $(1 - x)^{-n}$, where n is a positive integer, using the extended binomial theorem.

Solution: By the extended binomial theorem, it follows that

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

Using Example 8, which provides a simple formula for $\binom{-n}{k}$, we obtain

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

Replacing x by $-x$, we find that

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

◄

Table 1 presents a useful summary of some generating functions that arise frequently.

Remark: Note that the second and third formulae in this table can be deduced from the first formula by substituting ax and x^r for x, respectively. Similarly, the sixth and seventh formulae can be deduced from the fifth formula using the same substitutions. The tenth and eleventh can be deduced from the ninth formula by substituting $-x$ and ax for x, respectively. Also, some of the formulae in this table can be derived from other formulae using methods from calculus (such as differentiation and integration). Students are encouraged to know the core formulae in this table (that is, formulae from which the others can be derived, perhaps the first, fourth, fifth, eighth, ninth, twelfth, and thirteenth formulae) and understand how to derive the other formulae from these core formulae.

TABLE 1 Useful Generating Functions.

$G(x)$	a_k
$(1+x)^n = \sum_{k=0}^{n} C(n,k)x^k$ $= 1 + C(n,1)x + C(n,2)x^2 + \cdots + x^n$	$C(n,k)$
$(1+ax)^n = \sum_{k=0}^{n} C(n,k)a^k x^k$ $= 1 + C(n,1)ax + C(n,2)a^2x^2 + \cdots + a^n x^n$	$C(n,k)a^k$
$(1+x^r)^n = \sum_{k=0}^{n} C(n,k)x^{rk}$ $= 1 + C(n,1)x^r + C(n,2)x^{2r} + \cdots + x^{rn}$	$C(n,k/r)$ if $r \mid k$; 0 otherwise
$\frac{1-x^{n+1}}{1-x} = \sum_{k=0}^{n} x^k = 1 + x + x^2 + \cdots + x^n$	1 if $k \le n$; 0 otherwise
$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + \cdots$	1
$\frac{1}{1-ax} = \sum_{k=0}^{\infty} a^k x^k = 1 + ax + a^2x^2 + \cdots$	a^k
$\frac{1}{1-x^r} = \sum_{k=0}^{\infty} x^{rk} = 1 + x^r + x^{2r} + \cdots$	1 if $r \mid k$; 0 otherwise
$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} (k+1)x^k = 1 + 2x + 3x^2 + \cdots$	$k+1$
$\frac{1}{(1-x)^n} = \sum_{k=0}^{\infty} C(n+k-1,k)x^k$ $= 1 + C(n,1)x + C(n+1,2)x^2 + \cdots$	$C(n+k-1,k) = C(n+k-1,n-1)$
$\frac{1}{(1+x)^n} = \sum_{k=0}^{\infty} C(n+k-1,k)(-1)^k x^k$ $= 1 - C(n,1)x + C(n+1,2)x^2 - \cdots$	$(-1)^k C(n+k-1,k) = (-1)^k C(n+k-1,n-1)$
$\frac{1}{(1-ax)^n} = \sum_{k=0}^{\infty} C(n+k-1,k)a^k x^k$ $= 1 + C(n,1)ax + C(n+1,2)a^2x^2 + \cdots$	$C(n+k-1,k)a^k = C(n+k-1,n-1)a^k$
$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$	$1/k!$
$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$	$(-1)^{k+1}/k$

Note: The series for the last two generating functions can be found in most calculus books when power series are discussed.

8.4.3 Counting Problems and Generating Functions

Generating functions can be used to solve a wide variety of counting problems. In particular, they can be used to count the number of combinations of various types. In Chapter 6 we developed techniques to count the r-combinations from a set with n elements when repetition is allowed and additional constraints may exist. Such problems are equivalent to counting the solutions to equations of the form

$$e_1 + e_2 + \cdots + e_n = C,$$

where C is a constant and each e_i is a nonnegative integer that may be subject to a specified constraint. Generating functions can also be used to solve counting problems of this type, as Examples 10–12 show.

EXAMPLE 10 Find the number of solutions of

Extra Examples

$$e_1 + e_2 + e_3 = 17,$$

where e_1, e_2, and e_3 are nonnegative integers with $2 \le e_1 \le 5$, $3 \le e_2 \le 6$, and $4 \le e_3 \le 7$.

Solution: The number of solutions with the indicated constraints is the coefficient of x^{17} in the expansion of

$$(x^2 + x^3 + x^4 + x^5)(x^3 + x^4 + x^5 + x^6)(x^4 + x^5 + x^6 + x^7).$$

This follows because we obtain a term equal to x^{17} in the product by picking a term in the first sum x^{e_1}, a term in the second sum x^{e_2}, and a term in the third sum x^{e_3}, where the exponents e_1, e_2, and e_3 satisfy the equation $e_1 + e_2 + e_3 = 17$ and the given constraints.

It is not hard to see that the coefficient of x^{17} in this product is 3. Hence, there are three solutions. (Note that the calculating of this coefficient involves about as much work as enumerating all the solutions of the equation with the given constraints. However, the method that this illustrates often can be used to solve wide classes of counting problems with special formulae, as we will see. Furthermore, a computer algebra system can be used to do such computations.) ◀

EXAMPLE 11 In how many different ways can eight identical cookies be distributed among three distinct children if each child receives at least two cookies and no more than four cookies?

Solution: Because each child receives at least two but no more than four cookies, for each child there is a factor equal to

$$(x^2 + x^3 + x^4)$$

in the generating function for the sequence $\{c_n\}$, where c_n is the number of ways to distribute n cookies. Because there are three children, this generating function is

$$(x^2 + x^3 + x^4)^3.$$

We need the coefficient of x^8 in this product. The reason is that the x^8 terms in the expansion correspond to the ways that three terms can be selected, with one from each factor, that have exponents adding up to 8. Furthermore, the exponents of the term from the first, second, and third factors are the numbers of cookies the first, second, and third children receive, respectively. Computation shows that this coefficient equals 6. Hence, there are six ways to distribute the cookies so that each child receives at least two, but no more than four, cookies. ◀

EXAMPLE 12 Use generating functions to determine the number of ways to insert tokens worth \$1, \$2, and \$5 into a vending machine to pay for an item that costs r dollars in both the cases when the order in which the tokens are inserted does not matter and when the order does matter. (For example, there are two ways to pay for an item that costs \$3 when the order in which the tokens are inserted does not matter: inserting three \$1 tokens or one \$1 token and a \$2 token. When the order matters, there are three ways: inserting three \$1 tokens, inserting a \$1 token and then a \$2 token, or inserting a \$2 token and then a \$1 token.)

Solution: Consider the case when the order in which the tokens are inserted does not matter. Here, all we care about is the number of each token used to produce a total of r dollars. Because we can use any number of \$1 tokens, any number of \$2 tokens, and any number of \$5 tokens, the answer is the coefficient of x^r in the generating function

$$(1 + x + x^2 + x^3 + \cdots)(1 + x^2 + x^4 + x^6 + \cdots)(1 + x^5 + x^{10} + x^{15} + \cdots).$$

(The first factor in this product represents the \$1 tokens used, the second the \$2 tokens used, and the third the \$5 tokens used.) For example, the number of ways to pay for an item costing \$7 using \$1, \$2, and \$5 tokens is given by the coefficient of x^7 in this expansion, which equals 6.

When the order in which the tokens are inserted matters, the number of ways to insert exactly n tokens to produce a total of r dollars is the coefficient of x^r in

$$(x + x^2 + x^5)^n,$$

because each of the r tokens may be a \$1 token, a \$2 token, or a \$5 token. Because any number of tokens may be inserted, the number of ways to produce r dollars using \$1, \$2, or \$5 tokens, when the order in which the tokens are inserted matters, is the coefficient of x^r in

$$\begin{aligned} 1 + (x + x^2 + x^5) + (x + x^2 + x^5)^2 + \cdots &= \frac{1}{1 - (x + x^2 + x^5)} \\ &= \frac{1}{1 - x - x^2 - x^5}, \end{aligned}$$

where we have added the number of ways to insert 0 tokens, 1 token, 2 tokens, 3 tokens, and so on, and where we have used the identity $1/(1 - x) = 1 + x + x^2 + \cdots$ with x replaced with $x + x^2 + x^5$. For example, the number of ways to pay for an item costing \$7 using \$1, \$2, and \$5 tokens, when the order in which the tokens are used matters, is the coefficient of x^7 in this expansion, which equals 26. [*Hint:* To see that this coefficient equals 26 requires the addition of the coefficients of x^7 in the expansions $(x + x^2 + x^5)^k$ for $2 \leq k \leq 7$. This can be done by hand with considerable computation, or a computer algebra system can be used.] ◀

Example 13 shows the versatility of generating functions when used to solve problems with differing assumptions.

EXAMPLE 13 Use generating functions to find the number of k-combinations of a set with n elements. Assume that the binomial theorem has already been established.

Solution: Each of the n elements in the set contributes the term $(1 + x)$ to the generating function $f(x) = \sum_{k=0}^{n} a_k x^k$. Here $f(x)$ is the generating function for $\{a_k\}$, where a_k represents the number of k-combinations of a set with n elements. Hence,

$$f(x) = (1 + x)^n.$$

But by the binomial theorem, we have

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k,$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Hence, $C(n, k)$, the number of k-combinations of a set with n elements, is

$$\frac{n!}{k!(n-k)!}.$$

◀

Remark: We proved the binomial theorem in Section 6.4 using the formula for the number of r-combinations of a set with n elements. This example shows that the binomial theorem, which can be proved by mathematical induction, can be used to derive the formula for the number of r-combinations of a set with n elements.

EXAMPLE 14 Use generating functions to find the number of r-combinations from a set with n elements when repetition of elements is allowed.

Solution: Let $G(x)$ be the generating function for the sequence $\{a_r\}$, where a_r equals the number of r-combinations of a set with n elements with repetitions allowed. That is, $G(x) = \sum_{r=0}^{\infty} a_r x^r$. Because we can select any number of a particular member of the set with n elements when we form an r-combination with repetition allowed, each of the n elements contributes $(1 + x + x^2 + x^3 + \cdots)$ to a product expansion for $G(x)$. Each element contributes this factor because it may be selected zero times, one time, two times, three times, and so on, when an r-combination is formed (with a total of r elements selected). Because there are n elements in the set and each contributes this same factor to $G(x)$, we have

$$G(x) = (1 + x + x^2 + \cdots)^n.$$

As long as $|x| < 1$, we have $1 + x + x^2 + \cdots = 1/(1 - x)$, so

$$G(x) = 1/(1 - x)^n = (1 - x)^{-n}.$$

Applying the extended binomial theorem (Theorem 2), it follows that

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

The number of r-combinations of a set with n elements with repetitions allowed, when r is a positive integer, is the coefficient a_r of x^r in this sum. Consequently, using Example 8 we find that a_r equals

$$\begin{aligned}\binom{-n}{r}(-1)^r &= (-1)^r C(n + r - 1, r) \cdot (-1)^r \\ &= C(n + r - 1, r).\end{aligned}$$

◀

Note that the result in Example 14 is the same result we stated as Theorem 2 in Section 6.5.

EXAMPLE 15 Use generating functions to find the number of ways to select r objects of n different kinds if we must select at least one object of each kind.

Solution: Because we need to select at least one object of each kind, each of the n kinds of objects contributes the factor $(x + x^2 + x^3 + \cdots)$ to the generating function $G(x)$ for the sequence $\{a_r\}$, where a_r is the number of ways to select r objects of n different kinds if we need at least one object of each kind. Hence,

$$G(x) = (x + x^2 + x^3 + \cdots)^n = x^n(1 + x + x^2 + \cdots)^n = x^n/(1 - x)^n.$$

Using the extended binomial theorem and Example 8, we have

$$\begin{aligned} G(x) &= x^n/(1 - x)^n \\ &= x^n \cdot (1 - x)^{-n} \\ &= x^n \sum_{r=0}^{\infty} \binom{-n}{r} (-x)^r \\ &= x^n \sum_{r=0}^{\infty} (-1)^r C(n + r - 1, r)(-1)^r x^r \\ &= \sum_{r=0}^{\infty} C(n + r - 1, r)x^{n+r} \\ &= \sum_{t=n}^{\infty} C(t - 1, t - n)x^t \\ &= \sum_{r=n}^{\infty} C(r - 1, r - n)x^r. \end{aligned}$$

We have shifted the summation in the next-to-last equality by setting $t = n + r$ so that $t = n$ when $r = 0$ and $n + r - 1 = t - 1$, and then we replaced t by r as the index of summation in the last equality to return to our original notation. Hence, there are $C(r - 1, r - n)$ ways to select r objects of n different kinds if we must select at least one object of each kind. ◀

8.4.4 Using Generating Functions to Solve Recurrence Relations

We can find the solution to a recurrence relation and its initial conditions by finding an explicit formula for the associated generating function. This is illustrated in Examples 16 and 17.

EXAMPLE 16 Solve the recurrence relation $a_k = 3a_{k-1}$ for $k = 1, 2, 3, \ldots$ and initial condition $a_0 = 2$.

Extra Examples

Solution: Let $G(x)$ be the generating function for the sequence $\{a_k\}$, that is, $G(x) = \sum_{k=0}^{\infty} a_k x^k$. First note that

$$xG(x) = \sum_{k=0}^{\infty} a_k x^{k+1} = \sum_{k=1}^{\infty} a_{k-1} x^k.$$

Using the recurrence relation, we see that

$$\begin{aligned}
G(x) - 3xG(x) &= \sum_{k=0}^{\infty} a_k x^k - 3\sum_{k=1}^{\infty} a_{k-1}x^k \\
&= a_0 + \sum_{k=1}^{\infty}(a_k - 3a_{k-1})x^k \\
&= 2,
\end{aligned}$$

because $a_0 = 2$ and $a_k = 3a_{k-1}$. Thus,

$$G(x) - 3xG(x) = (1 - 3x)G(x) = 2.$$

Solving for $G(x)$ shows that $G(x) = 2/(1 - 3x)$. Using the identity $1/(1 - ax) = \sum_{k=0}^{\infty} a^k x^k$, from Table 1, we have

$$G(x) = 2\sum_{k=0}^{\infty} 3^k x^k = \sum_{k=0}^{\infty} 2 \cdot 3^k x^k.$$

Consequently, $a_k = 2 \cdot 3^k$. ◀

EXAMPLE 17 Suppose that a valid codeword is an n-digit number in decimal notation containing an even number of 0s. Let a_n denote the number of valid codewords of length n. In Example 4 of Section 8.1 we showed that the sequence $\{a_n\}$ satisfies the recurrence relation

$$a_n = 8a_{n-1} + 10^{n-1}$$

and the initial condition $a_1 = 9$. Use generating functions to find an explicit formula for a_n.

Solution: To make our work with generating functions simpler, we extend this sequence by setting $a_0 = 1$; when we assign this value to a_0 and use the recurrence relation, we have $a_1 = 8a_0 + 10^0 = 8 + 1 = 9$, which is consistent with our original initial condition. (It also makes sense because there is one code word of length 0—the empty string.)

We multiply both sides of the recurrence relation by x^n to obtain

$$a_n x^n = 8a_{n-1}x^n + 10^{n-1}x^n.$$

Let $G(x) = \sum_{n=0}^{\infty} a_n x^n$ be the generating function of the sequence $a_0, a_1, a_2, \ldots$. We sum both sides of the last equation starting with $n = 1$, to find that

$$\begin{aligned}
G(x) - 1 = \sum_{n=1}^{\infty} a_n x^n &= \sum_{n=1}^{\infty}(8a_{n-1}x^n + 10^{n-1}x^n) \\
&= 8\sum_{n=1}^{\infty} a_{n-1}x^n + \sum_{n=1}^{\infty} 10^{n-1}x^n \\
&= 8x\sum_{n=1}^{\infty} a_{n-1}x^{n-1} + x\sum_{n=1}^{\infty} 10^{n-1}x^{n-1} \\
&= 8x\sum_{n=0}^{\infty} a_n x^n + x\sum_{n=0}^{\infty} 10^n x^n \\
&= 8xG(x) + x/(1 - 10x),
\end{aligned}$$

where we have used Example 5 to evaluate the second summation. Therefore, we have

$$G(x) - 1 = 8xG(x) + x/(1 - 10x).$$

Solving for $G(x)$ shows that

$$G(x) = \frac{1 - 9x}{(1 - 8x)(1 - 10x)}.$$

Expanding the right-hand side of this equation into partial fractions (as is done in the integration of rational functions studied in calculus) gives

$$G(x) = \frac{1}{2}\left(\frac{1}{1 - 8x} + \frac{1}{1 - 10x}\right).$$

Using Example 5 twice (once with $a = 8$ and once with $a = 10$) gives

$$G(x) = \frac{1}{2}\left(\sum_{n=0}^{\infty} 8^n x^n + \sum_{n=0}^{\infty} 10^n x^n\right)$$

$$= \sum_{n=0}^{\infty} \frac{1}{2}(8^n + 10^n)x^n.$$

Consequently, we have shown that

$$a_n = \frac{1}{2}(8^n + 10^n).$$

◀

8.4.5 Proving Identities via Generating Functions

In Chapter 6 we saw how combinatorial identities could be established using combinatorial proofs. Here we will show that such identities, as well as identities for extended binomial coefficients, can be proved using generating functions. Sometimes the generating function approach is simpler than other approaches, especially when it is simpler to work with the closed form of a generating function than with the terms of the sequence themselves. We illustrate how generating functions can be used to prove identities with Example 18.

EXAMPLE 18 Use generating functions to show that

$$\sum_{k=0}^{n} C(n, k)^2 = C(2n, n)$$

whenever n is a positive integer.

Solution: First note that by the binomial theorem $C(2n, n)$ is the coefficient of x^n in $(1 + x)^{2n}$. However, we also have

$$(1 + x)^{2n} = [(1 + x)^n]^2$$

$$= [C(n, 0) + C(n, 1)x + C(n, 2)x^2 + \cdots + C(n, n)x^n]^2.$$

The coefficient of x^n in this expression is

$$C(n,0)C(n,n) + C(n,1)C(n,n-1) + C(n,2)C(n,n-2) + \cdots + C(n,n)C(n,0).$$

This equals $\sum_{k=0}^{n} C(n,k)^2$, because $C(n,n-k) = C(n,k)$. Because both $C(2n,n)$ and $\sum_{k=0}^{n} C(n,k)^2$ represent the coefficient of x^n in $(1+x)^{2n}$, they must be equal. ◀

Exercises 44 and 45 ask that Pascal's identity and Vandermonde's identity be proved using generating functions.

Exercises

1. Find the generating function for the finite sequence 2, 2, 2, 2, 2, 2.

2. Find the generating function for the finite sequence 1, 4, 16, 64, 256.

In Exercises 3–8, by a **closed form** we mean an algebraic expression not involving a summation over a range of values or the use of ellipses.

3. Find a closed form for the generating function for each of these sequences. (For each sequence, use the most obvious choice of a sequence that follows the pattern of the initial terms listed.)

a) 0, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, …
b) 0, 0, 0, 1, 1, 1, 1, 1, 1, …
c) 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, …
d) 2, 4, 8, 16, 32, 64, 128, 256, …
e) $\binom{7}{0}, \binom{7}{1}, \binom{7}{2}, \ldots, \binom{7}{7}$, 0, 0, 0, 0, 0, …
f) 2, −2, 2, −2, 2, −2, 2, −2, …
g) 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, …
h) 0, 0, 0, 1, 2, 3, 4, …

4. Find a closed form for the generating function for each of these sequences. (Assume a general form for the terms of the sequence, using the most obvious choice of such a sequence.)

a) −1, −1, −1, −1, −1, −1, −1, 0, 0, 0, 0, 0, 0, …
b) 1, 3, 9, 27, 81, 243, 729, …
c) 0, 0, 3, −3, 3, −3, 3, −3, …
d) 1, 2, 1, 1, 1, 1, 1, 1, 1, …
e) $\binom{7}{0}, 2\binom{7}{1}, 2^2\binom{7}{2}, \ldots, 2^7\binom{7}{7}$, 0, 0, 0, 0, …
f) −3, 3, −3, 3, −3, 3, …
g) 0, 1, −2, 4, −8, 16, −32, 64, …
h) 1, 0, 1, 0, 1, 0, 1, 0, …

5. Find a closed form for the generating function for the sequence $\{a_n\}$, where

a) $a_n = 5$ for all $n = 0, 1, 2, \ldots$.
b) $a_n = 3^n$ for all $n = 0, 1, 2, \ldots$.
c) $a_n = 2$ for $n = 3, 4, 5, \ldots$ and $a_0 = a_1 = a_2 = 0$.
d) $a_n = 2n + 3$ for all $n = 0, 1, 2, \ldots$.
e) $a_n = \binom{8}{n}$ for all $n = 0, 1, 2, \ldots$.
f) $a_n = \binom{n+4}{n}$ for all $n = 0, 1, 2, \ldots$.

6. Find a closed form for the generating function for the sequence $\{a_n\}$, where

a) $a_n = -1$ for all $n = 0, 1, 2, \ldots$.
b) $a_n = 2^n$ for $n = 1, 2, 3, 4, \ldots$ and $a_0 = 0$.
c) $a_n = n - 1$ for $n = 0, 1, 2, \ldots$.
d) $a_n = 1/(n+1)!$ for $n = 0, 1, 2, \ldots$.
e) $a_n = \binom{n}{2}$ for $n = 0, 1, 2, \ldots$.
f) $a_n = \binom{10}{n+1}$ for $n = 0, 1, 2, \ldots$.

7. For each of these generating functions, provide a closed formula for the sequence it determines.

a) $(3x-4)^3$
b) $(x^3+1)^3$
c) $1/(1-5x)$
d) $x^3/(1+3x)$
e) $x^2 + 3x + 7 + (1/(1-x^2))$
f) $(x^4/(1-x^4)) - x^3 - x^2 - x - 1$
g) $x^2/(1-x)^2$
h) $2e^{2x}$

8. For each of these generating functions, provide a closed formula for the sequence it determines.

a) $(x^2+1)^3$
b) $(3x-1)^3$
c) $1/(1-2x^2)$
d) $x^2/(1-x)^3$
e) $x - 1 + (1/(1-3x))$
f) $(1+x^3)/(1+x)^3$
***g)** $x/(1+x+x^2)$
h) $e^{3x^2} - 1$

9. Find the coefficient of x^{10} in the power series of each of these functions.

a) $(1 + x^5 + x^{10} + x^{15} + \cdots)^3$
b) $(x^3 + x^4 + x^5 + x^6 + x^7 + \cdots)^3$
c) $(x^4 + x^5 + x^6)(x^3 + x^4 + x^5 + x^6 + x^7)(1 + x + x^2 + x^3 + x^4 + \cdots)$
d) $(x^2 + x^4 + x^6 + x^8 + \cdots)(x^3 + x^6 + x^9 + \cdots)(x^4 + x^8 + x^{12} + \cdots)$
e) $(1 + x^2 + x^4 + x^6 + x^8 + \cdots)(1 + x^4 + x^8 + x^{12} + \cdots)(1 + x^6 + x^{12} + x^{18} + \cdots)$

10. Find the coefficient of x^9 in the power series of each of these functions.

a) $(1 + x^3 + x^6 + x^9 + \cdots)^3$
b) $(x^2 + x^3 + x^4 + x^5 + x^6 + \cdots)^3$
c) $(x^3 + x^5 + x^6)(x^3 + x^4)(x + x^2 + x^3 + x^4 + \cdots)$
d) $(x + x^4 + x^7 + x^{10} + \cdots)(x^2 + x^4 + x^6 + x^8 + \cdots)$
e) $(1 + x + x^2)^3$

11. Find the coefficient of x^{10} in the power series of each of these functions.

a) $1/(1-2x)$
b) $1/(1+x)^2$
c) $1/(1-x)^3$
d) $1/(1+2x)^4$
e) $x^4/(1-3x)^3$

12. Find the coefficient of x^{12} in the power series of each of these functions.

a) $1/(1+3x)$
b) $1/(1-2x)^2$
c) $1/(1+x)^8$
d) $1/(1-4x)^3$
e) $x^3/(1+4x)^2$

13. Use generating functions to determine the number of different ways 10 identical balloons can be given to four children if each child receives at least two balloons.

14. Use generating functions to determine the number of different ways 12 identical action figures can be given to five children so that each child receives at most three action figures.

15. Use generating functions to determine the number of different ways 15 identical stuffed animals can be given to six children so that each child receives at least one but no more than three stuffed animals.

16. Use generating functions to find the number of ways to choose a dozen bagels from three varieties—egg, salty, and plain—if at least two bagels of each kind but no more than three salty bagels are chosen.

17. In how many ways can 25 identical donuts be distributed to four police officers so that each officer gets at least three but no more than seven donuts?

18. Use generating functions to find the number of ways to select 14 balls from a jar containing 100 red balls, 100 blue balls, and 100 green balls so that no fewer than 3 and no more than 10 blue balls are selected. Assume that the order in which the balls are drawn does not matter.

19. What is the generating function for the sequence $\{c_k\}$, where c_k is the number of ways to make change for k dollars using \$1 bills, \$2 bills, \$5 bills, and \$10 bills?

20. What is the generating function for the sequence $\{c_k\}$, where c_k represents the number of ways to make change for k pesos using bills worth 10 pesos, 20 pesos, 50 pesos, and 100 pesos?

21. Give a combinatorial interpretation of the coefficient of x^4 in the expansion $(1+x+x^2+x^3+\cdots)^3$. Use this interpretation to find this number.

22. Give a combinatorial interpretation of the coefficient of x^6 in the expansion $(1+x+x^2+x^3+\cdots)^n$. Use this interpretation to find this number.

23. a) What is the generating function for $\{a_k\}$, where a_k is the number of solutions of $x_1+x_2+x_3=k$ when x_1, x_2, and x_3 are integers with $x_1 \ge 2$, $0 \le x_2 \le 3$, and $2 \le x_3 \le 5$?
b) Use your answer to part (a) to find a_6.

24. a) What is the generating function for $\{a_k\}$, where a_k is the number of solutions of $x_1+x_2+x_3+x_4=k$ when x_1, x_2, x_3, and x_4 are integers with $x_1 \ge 3$, $1 \le x_2 \le 5$, $0 \le x_3 \le 4$, and $x_4 \ge 1$?
b) Use your answer to part (a) to find a_7.

25. Explain how generating functions can be used to find the number of ways in which postage of r cents can be pasted on an envelope using 3-cent, 4-cent, and 20-cent stamps.

a) Assume that the order the stamps are pasted on does not matter.
b) Assume that the order in which the stamps are pasted on matters.
c) Use your answer to part (a) to determine the number of ways 46 cents of postage can be pasted on an envelope using 3-cent, 4-cent, and 20-cent stamps when the order the stamps are pasted on does not matter. (Use of a computer algebra program is advised.)
d) Use your answer to part (b) to determine the number of ways 46 cents of postage can be pasted in a row on an envelope using 3-cent, 4-cent, and 20-cent stamps when the order in which the stamps are pasted on matters. (Use of a computer algebra program is advised.)

26. Explain how generating functions can be used to find the number of ways in which postage of r cents can be pasted on an envelope using 2-cent, 7-cent, 13-cent, and 32-cent stamps.

a) Assume that the order the stamps are pasted on does not matter.
b) Assume that the order the stamps are pasted on matters.
c) Use your answer to part (a) to determine the number of ways 49 cents of postage can be pasted on an envelope using 2-cent, 7-cent, 13-cent, and 32-cent stamps when the order the stamps are pasted on does not matter. (Use of a computer algebra program is advised.)
d) Use your answer to part (b) to determine the number of ways 49 cents of postage can be pasted on an envelope using 2-cent, 7-cent, 13-cent, and 32-cent stamps when the order in which the stamps are pasted on matters. (Use of a computer algebra program is advised.)

27. Customers at a quirky tropical fruit stand can buy at most four mangos, at most two passion fruit, any even number of papayas, three or more coconuts, and carambolas in groups of five.

a) Explain how generating functions can be used to find the number of ways a customer can buy n pieces of these fruits, following the restrictions listed.
b) Use your answer in part (a) to determine the number of ways you can buy a dozen pieces of these fruits.

28. a) Show that $1/(1-x-x^2-x^3-x^4-x^5-x^6)$ is the generating function for the number of ways that the sum n can be obtained when a die is rolled repeatedly and the order of the rolls matters.
b) Use part (a) to find the number of ways to roll a total of 8 when a die is rolled repeatedly, and the order of the rolls matters. (Use of a computer algebra package is advised.)

29. Use generating functions (and a computer algebra package, if available) to find the number of ways to make change for \$1 using

a) dimes and quarters.
b) nickels, dimes, and quarters.
c) pennies, dimes, and quarters.
d) pennies, nickels, dimes, and quarters.

30. Use generating functions (and a computer algebra package, if available) to find the number of ways to make change for \$1 using pennies, nickels, dimes, and quarters with

a) no more than 10 pennies.
b) no more than 10 pennies and no more than 10 nickels.
***c)** no more than 10 coins.

31. Use generating functions to find the number of ways to make change for \$100 using

a) \$10, \$20, and \$50 bills.
b) \$5, \$10, \$20, and \$50 bills.
c) \$5, \$10, \$20, and \$50 bills if at least one bill of each denomination is used.
d) \$5, \$10, and \$20 bills if at least one and no more than four of each denomination is used.

32. If $G(x)$ is the generating function for the sequence $\{a_k\}$, what is the generating function for each of these sequences?

a) $2a_0, 2a_1, 2a_2, 2a_3, \ldots$
b) $0, a_0, a_1, a_2, a_3, \ldots$ (assuming that terms follow the pattern of all but the first term)
c) $0, 0, 0, 0, a_2, a_3, \ldots$ (assuming that terms follow the pattern of all but the first four terms)
d) $a_2, a_3, a_4, \ldots$
e) $a_1, 2a_2, 3a_3, 4a_4, \ldots$ [*Hint: Calculus required* here.]
f) $a_0^2, \quad 2a_0a_1, \quad a_1^2 + 2a_0a_2, \quad 2a_0a_3 + 2a_1a_2, \quad 2a_0a_4 + 2a_1a_3 + a_2^2, \ldots$

33. If $G(x)$ is the generating function for the sequence $\{a_k\}$, what is the generating function for each of these sequences?

a) $0, 0, 0, a_3, a_4, a_5, \ldots$ (assuming that terms follow the pattern of all but the first three terms)
b) $a_0, 0, a_1, 0, a_2, 0, \ldots$
c) $0, 0, 0, 0, a_0, a_1, a_2, \ldots$ (assuming that terms follow the pattern of all but the first four terms)
d) $a_0, 2a_1, 4a_2, 8a_3, 16a_4, \ldots$
e) $0, a_0, a_1/2, a_2/3, a_3/4, \ldots$ [*Hint: Calculus required* here.]
f) $a_0, a_0 + a_1, a_0 + a_1 + a_2, a_0 + a_1 + a_2 + a_3, \ldots$

34. Use generating functions to solve the recurrence relation $a_k = 7a_{k-1}$ with the initial condition $a_0 = 5$.

35. Use generating functions to solve the recurrence relation $a_k = 3a_{k-1} + 2$ with the initial condition $a_0 = 1$.

36. Use generating functions to solve the recurrence relation $a_k = 3a_{k-1} + 4^{k-1}$ with the initial condition $a_0 = 1$.

37. Use generating functions to solve the recurrence relation $a_k = 5a_{k-1} - 6a_{k-2}$ with initial conditions $a_0 = 6$ and $a_1 = 30$.

38. Use generating functions to solve the recurrence relation $a_k = a_{k-1} + 2a_{k-2} + 2^k$ with initial conditions $a_0 = 4$ and $a_1 = 12$.

39. Use generating functions to solve the recurrence relation $a_k = 4a_{k-1} - 4a_{k-2} + k^2$ with initial conditions $a_0 = 2$ and $a_1 = 5$.

40. Use generating functions to solve the recurrence relation $a_k = 2a_{k-1} + 3a_{k-2} + 4^k + 6$ with initial conditions $a_0 = 20$, $a_1 = 60$.

41. Use generating functions to find an explicit formula for the Fibonacci numbers.

***42. a)** Show that if n is a positive integer, then

$$\binom{-1/2}{n} = \binom{2n}{n} \Big/ (-4)^n.$$

b) Use the extended binomial theorem and part (a) to show that the coefficient of x^n in the expansion of $(1-4x)^{-1/2}$ is $\binom{2n}{n}$ for all nonnegative integers n.

***43.** (*Calculus required*) Let $\{C_n\}$ be the sequence of Catalan numbers, that is, the solution to the recurrence relation $C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1}$ with $C_0 = C_1 = 1$ (see Example 5 in Section 8.1).

a) Show that if $G(x)$ is the generating function for the sequence of Catalan numbers, then $xG(x)^2 - G(x) + 1 = 0$. Conclude (using the initial conditions) that $G(x) = (1 - \sqrt{1-4x})/(2x)$.

b) Use Exercise 42 to conclude that

$$G(x) = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} x^n,$$

so that

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

c) Show that $C_n \geq 2^{n-1}$ for all positive integers n.

44. Use generating functions to prove Pascal's identity: $C(n, r) = C(n-1, r) + C(n-1, r-1)$ when n and r are positive integers with $r < n$. [*Hint:* Use the identity $(1+x)^n = (1+x)^{n-1} + x(1+x)^{n-1}$.]

45. Use generating functions to prove Vandermonde's identity: $C(m+n, r) = \sum_{k=0}^{r} C(m, r-k)C(n, k)$, whenever m, n, and r are nonnegative integers with r not exceeding either m or n. [*Hint:* Look at the coefficient of x^r in both sides of $(1+x)^{m+n} = (1+x)^m(1+x)^n$.]

46. This exercise shows how to use generating functions to derive a formula for the sum of the first n squares.

a) Show that $(x^2 + x)/(1-x)^4$ is the generating function for the sequence $\{a_n\}$, where $a_n = 1^2 + 2^2 + \cdots + n^2$.

b) Use part (a) to find an explicit formula for the sum $1^2 + 2^2 + \cdots + n^2$.

The **exponential generating function** for the sequence $\{a_n\}$ is the series

$$\sum_{n=0}^{\infty} \frac{a_n}{n!} x^n.$$

For example, the exponential generating function for the sequence $1, 1, 1, \ldots$ is the function $\sum_{n=0}^{\infty} x^n/n! = e^x$. (You will find this particular series useful in these exercises.) Note that e^x is the (ordinary) generating function for the sequence $1, 1, 1/2!, 1/3!, 1/4!, \ldots$.

47. Find a closed form for the exponential generating function for the sequence $\{a_n\}$, where

a) $a_n = 2$. **b)** $a_n = (-1)^n$.
c) $a_n = 3^n$. **d)** $a_n = n + 1$.
e) $a_n = 1/(n + 1)$.

48. Find a closed form for the exponential generating function for the sequence $\{a_n\}$, where

a) $a_n = (-2)^n$. **b)** $a_n = -1$.
c) $a_n = n$. **d)** $a_n = n(n - 1)$.
e) $a_n = 1/((n + 1)(n + 2))$.

49. Find the sequence with each of these functions as its exponential generating function.

a) $f(x) = e^{-x}$ **b)** $f(x) = 3x^{2x}$
c) $f(x) = e^{3x} - 3e^{2x}$ **d)** $f(x) = (1 - x) + e^{-2x}$
e) $f(x) = e^{-2x} - (1/(1 - x))$
f) $f(x) = e^{-3x} - (1 + x) + (1/(1 - 2x))$
g) $f(x) = e^{x^2}$

50. Find the sequence with each of these functions as its exponential generating function.

a) $f(x) = e^{3x}$ **b)** $f(x) = 2e^{-3x+1}$
c) $f(x) = e^{4x} + e^{-4x}$ **d)** $f(x) = (1 + 2x) + e^{3x}$
e) $f(x) = e^x - (1/(1 + x))$
f) $f(x) = xe^x$ **g)** $f(x) = e^{x^3}$

51. A coding system encodes messages using strings of octal (base 8) digits. A codeword is considered valid if and only if it contains an even number of 7s.

a) Find a linear nonhomogeneous recurrence relation for the number of valid codewords of length n. What are the initial conditions?
b) Solve this recurrence relation using Theorem 6 in Section 8.2.
c) Solve this recurrence relation using generating functions.

***52.** A coding system encodes messages using strings of base 4 digits (that is, digits from the set $\{0, 1, 2, 3\}$). A codeword is valid if and only if it contains an even number of 0s and an even number of 1s. Let a_n equal the number of valid codewords of length n. Furthermore, let b_n, c_n, and d_n equal the number of strings of base 4 digits of length n with an even number of 0s and an odd number of 1s, with an odd number of 0s and an even number of 1s, and with an odd number of 0s and an odd number of 1s, respectively.

a) Show that $d_n = 4^n - a_n - b_n - c_n$. Use this to show that $a_{n+1} = 2a_n + b_n + c_n$, $b_{n+1} = b_n - c_n + 4^n$, and $c_{n+1} = c_n - b_n + 4^n$.
b) What are a_1, b_1, c_1, and d_1?
c) Use parts (a) and (b) to find a_3, b_3, c_3, and d_3.
d) Use the recurrence relations in part (a), together with the initial conditions in part (b), to set up three equations relating the generating functions $A(x)$, $B(x)$, and $C(x)$ for the sequences $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$, respectively.
e) Solve the system of equations from part (d) to get explicit formulae for $A(x)$, $B(x)$, and $C(x)$ and use these to get explicit formulae for a_n, b_n, c_n, and d_n.

Generating functions are useful in studying the number of different types of partitions of an integer n. A **partition** of a positive integer is a way to write this integer as the sum of positive integers where repetition is allowed and the order of the integers in the sum does not matter. For example, the partitions of 5 (with no restrictions) are $1 + 1 + 1 + 1 + 1$, $1 + 1 + 1 + 2$, $1 + 1 + 3$, $1 + 2 + 2$, $1 + 4$, $2 + 3$, and 5. Exercises 53–58 illustrate some of these uses.

53. Show that the coefficient $p(n)$ of x^n in the formal power series expansion of $1/((1-x)(1-x^2)(1-x^3)\cdots)$ equals the number of partitions of n.

54. Show that the coefficient $p_o(n)$ of x^n in the formal power series expansion of $1/((1-x)(1-x^3)(1-x^5)\cdots)$ equals the number of partitions of n into odd integers, that is, the number of ways to write n as the sum of odd positive integers, where the order does not matter and repetitions are allowed.

55. Show that the coefficient $p_d(n)$ of x^n in the formal power series expansion of $(1 + x)(1 + x^2)(1 + x^3)\cdots$ equals the number of partitions of n into distinct parts, that is, the number of ways to write n as the sum of positive integers, where the order does not matter but no repetitions are allowed.

56. Find $p_o(n)$, the number of partitions of n into odd parts with repetitions allowed, and $p_d(n)$, the number of partitions of n into distinct parts, for $1 \le n \le 8$, by writing each partition of each type for each integer.

57. Show that if n is a positive integer, then the number of partitions of n into distinct parts equals the number of partitions of n into odd parts with repetitions allowed; that is, $p_o(n) = p_d(n)$. [*Hint:* Show that the generating functions for $p_o(n)$ and $p_d(n)$ are equal.]

****58.** (*Requires calculus*) Use the generating function of $p(n)$ to show that $p(n) \le e^{C\sqrt{n}}$ for some constant C. [Hardy and Ramanujan showed that $p(n) \sim e^{\pi\sqrt{2/3}\sqrt{n}}/(4\sqrt{3}n)$, which means that the ratio of $p(n)$ and the right-hand side approaches 1 as n approaches infinity.]

Suppose that X is a random variable on a sample space S such that $X(s)$ is a nonnegative integer for all $s \in S$. The **probability generating function** for X is

Links

$$G_X(x) = \sum_{k=0}^{\infty} p(X(s) = k)x^k.$$

59. (*Requires calculus*) Show that if G_X is the probability generating function for a random variable X such that $X(s)$ is a nonnegative integer for all $s \in S$, then

a) $G_X(1) = 1$. **b)** $E(X) = G'_X(1)$.
c) $V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$.

60. Let X be the random variable whose value is n if the first success occurs on the nth trial when independent Bernoulli trials are performed, each with probability of success p.

a) Find a closed formula for the probability generating function G_X.
b) Find the expected value and the variance of X using Exercise 59 and the closed form for the probability generating function found in part (a).

61. Let m be a positive integer. Let X_m be the random variable whose value is n if the mth success occurs on the $(n+m)$th trial when independent Bernoulli trials are performed, each with probability of success p.

a) Using Exercise 32 in the Supplementary Exercises of Chapter 7, show that the probability generating function G_{X_m} is given by $G_{X_m}(x) = p^m/(1-qx)^m$, where $q = 1-p$.

b) Find the expected value and the variance of X_m using Exercise 59 and the closed form for the probability generating function in part (a).

62. Show that if X and Y are independent random variables on a sample space S such that $X(s)$ and $Y(s)$ are nonnegative integers for all $s \in S$, then $G_{X+Y}(x) = G_X(x)G_Y(x)$.

8.5 Inclusion–Exclusion

8.5.1 Introduction

A discrete mathematics class contains 30 women and 50 sophomores. How many students in the class are either women or sophomores? This question cannot be answered unless more information is provided. Adding the number of women in the class and the number of sophomores probably does not give the correct answer, because women sophomores are counted twice. This observation shows that the number of students in the class that are either sophomores or women is the sum of the number of women and the number of sophomores in the class minus the number of women sophomores. A technique for solving such counting problems was introduced in Section 6.1. In this section we will generalize the ideas introduced in that section to solve problems that require us to count the number of elements in the union of more than two sets.

8.5.2 The Principle of Inclusion–Exclusion

How many elements are in the union of two finite sets? In Section 2.2 we showed that the number of elements in the union of the two sets A and B is the sum of the numbers of elements in the sets minus the number of elements in their intersection. That is,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

As we showed in Section 6.1, the formula for the number of elements in the union of two sets is useful in counting problems. Examples 1–3 provide additional illustrations of the usefulness of this formula.

EXAMPLE 1 In a discrete mathematics class every student is a major in computer science or mathematics, or both. The number of students having computer science as a major (possibly along with mathematics) is 25; the number of students having mathematics as a major (possibly along with computer science) is 13; and the number of students majoring in both computer science and mathematics is 8. How many students are in this class?

Solution: Let A be the set of students in the class majoring in computer science and B be the set of students in the class majoring in mathematics. Then $A \cap B$ is the set of students in the class who are joint mathematics and computer science majors. Because every student in the class is majoring in either computer science or mathematics (or both), it follows that the number of students in the class is $|A \cup B|$. Therefore,

$$\begin{aligned} |A \cup B| &= |A| + |B| - |A \cap B| \\ &= 25 + 13 - 8 = 30. \end{aligned}$$

Therefore, there are 30 students in the class. This computation is illustrated in Figure 1. ◀

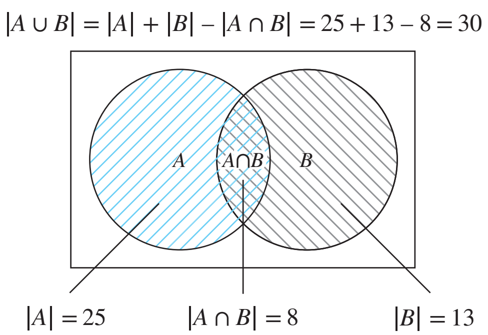


FIGURE 1 The set of students in a discrete mathematics class.

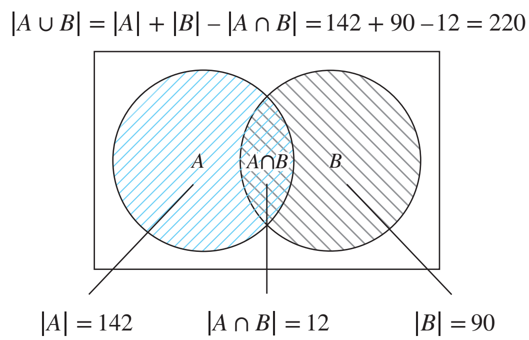


FIGURE 2 The set of positive integers not exceeding 1000 divisible by either 7 or 11.

EXAMPLE 2 How many positive integers not exceeding 1000 are divisible by 7 or 11?

Solution: Let A be the set of positive integers not exceeding 1000 that are divisible by 7, and let B be the set of positive integers not exceeding 1000 that are divisible by 11. Then $A \cup B$ is the set of integers not exceeding 1000 that are divisible by either 7 or 11, and $A \cap B$ is the set of integers not exceeding 1000 that are divisible by both 7 and 11. From Example 2 of Section 4.1, we know that among the positive integers not exceeding 1000 there are $\lfloor 1000/7 \rfloor$ integers divisible by 7 and $\lfloor 1000/11 \rfloor$ divisible by 11. Because 7 and 11 are relatively prime, the integers divisible by both 7 and 11 are those divisible by $7 \cdot 11$. Consequently, there are $\lfloor 1000/(11 \cdot 7) \rfloor$ positive integers not exceeding 1000 that are divisible by both 7 and 11. It follows that there are

$$\begin{aligned}|A \cup B| &= |A| + |B| - |A \cap B| \\ &= \left\lfloor \frac{1000}{7} \right\rfloor + \left\lfloor \frac{1000}{11} \right\rfloor - \left\lfloor \frac{1000}{7 \cdot 11} \right\rfloor \\ &= 142 + 90 - 12 = 220\end{aligned}$$

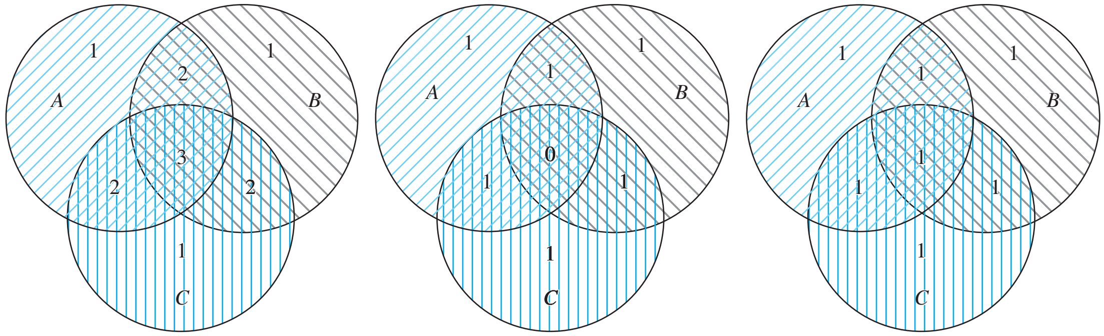
positive integers not exceeding 1000 that are divisible by either 7 or 11. This computation is illustrated in Figure 2. ◀

Example 3 shows how to find the number of elements in a finite universal set that are outside the union of two sets.

EXAMPLE 3 Suppose that there are 1807 freshmen at your school. Of these, 453 are taking a course in computer science, 567 are taking a course in mathematics, and 299 are taking courses in both computer science and mathematics. How many are not taking a course either in computer science or in mathematics?

Solution: To find the number of freshmen who are not taking a course in either mathematics or computer science, subtract the number that are taking a course in either of these subjects from the total number of freshmen. Let A be the set of all freshmen taking a course in computer science, and let B be the set of all freshmen taking a course in mathematics. It follows that $|A| = 453$, $|B| = 567$, and $|A \cap B| = 299$. The number of freshmen taking a course in either computer science or mathematics is

$$|A \cup B| = |A| + |B| - |A \cap B| = 453 + 567 - 299 = 721.$$



(a) Count of elements by $|A| + |B| + |C|$

(b) Count of elements by $|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|$

(c) Count of elements by $|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$

FIGURE 3 Finding a formula for the number of elements in the union of three sets.

Consequently, there are $1807 - 721 = 1086$ freshmen who are not taking a course in computer science or mathematics. ◀

We will now begin our development of a formula for the number of elements in the union of a finite number of sets. The formula we will develop is called the **principle of inclusion–exclusion**. For concreteness, before we consider unions of n sets, where n is any positive integer, we will derive a formula for the number of elements in the union of three sets A, B, and C. To construct this formula, we note that $|A| + |B| + |C|$ counts each element that is in exactly one of the three sets once, elements that are in exactly two of the sets twice, and elements in all three sets three times. This is illustrated in the first panel in Figure 3.

To remove the overcount of elements in more than one of the sets, we subtract the number of elements in the intersections of all pairs of the three sets. We obtain

$$|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C|.$$

This expression still counts elements that occur in exactly one of the sets once. An element that occurs in exactly two of the sets is also counted exactly once, because this element will occur in one of the three intersections of sets taken two at a time. However, those elements that occur in all three sets will be counted zero times by this expression, because they occur in all three intersections of sets taken two at a time. This is illustrated in the second panel in Figure 3.

To remedy this undercount, we add the number of elements in the intersection of all three sets. This final expression counts each element once, whether it is in one, two, or three of the sets. Thus,

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

This formula is illustrated in the third panel of Figure 3.

Example 4 illustrates how this formula can be used.

EXAMPLE 4 A total of 1232 students have taken a course in Spanish, 879 have taken a course in French, and 114 have taken a course in Russian. Further, 103 have taken courses in both Spanish and French, 23 have taken courses in both Spanish and Russian, and 14 have taken courses in both

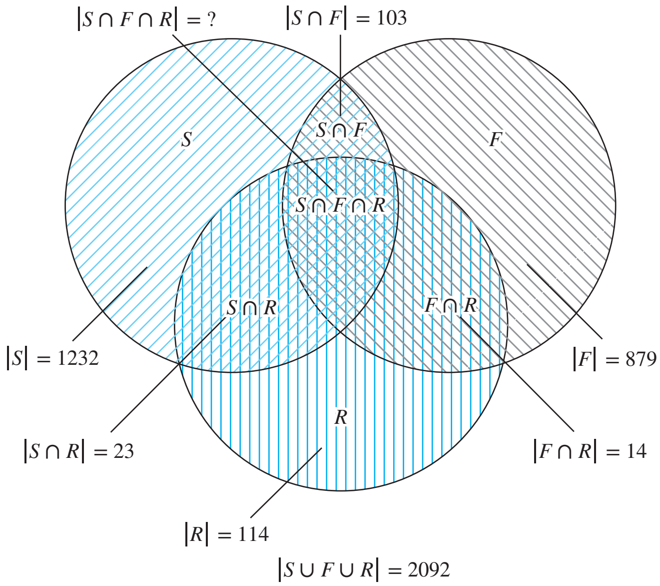


FIGURE 4 The set of students who have taken courses in Spanish, French, and Russian.

French and Russian. If 2092 students have taken at least one of Spanish, French, and Russian, how many students have taken a course in all three languages?

Solution: Let S be the set of students who have taken a course in Spanish, F the set of students who have taken a course in French, and R the set of students who have taken a course in Russian. Then

$$|S| = 1232, \quad |F| = 879, \quad |R| = 114,$$
$$|S \cap F| = 103, \ |S \cap R| = 23, \ |F \cap R| = 14,$$

and

$$|S \cup F \cup R| = 2092.$$

When we insert these quantities into the equation

$$|S \cup F \cup R| = |S| + |F| + |R| - |S \cap F| - |S \cap R| - |F \cap R| + |S \cap F \cap R|$$

we obtain

$$2092 = 1232 + 879 + 114 - 103 - 23 - 14 + |S \cap F \cap R|.$$

We now solve for $|S \cap F \cap R|$. We find that $|S \cap F \cap R| = 7$. Therefore, there are seven students who have taken courses in Spanish, French, and Russian. This is illustrated in Figure 4. ◀

We will now state and prove the **inclusion–exclusion principle** for n sets, where n is a positive integer. This priniciple tells us that we can count the elements in a union of n sets by adding the number of elements in the sets, then subtracting the sum of the number of elements in all intersections of two of these sets, then adding the number of elements in all intersections of three of these sets, and so on, until we reach the number of elements in the intersection of all the sets. It is added when there is an odd number of sets and added when there is an even number of sets.

THEOREM 1 **THE PRINCIPLE OF INCLUSION–EXCLUSION** Let $A_1, A_2, \ldots, A_n$ be finite sets. Then

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{1 \le i \le n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1} |A_1 \cap A_2 \cap \cdots \cap A_n|.$$

Proof: We will prove the formula by showing that an element in the union is counted exactly once by the right-hand side of the equation. Suppose that a is a member of exactly r of the sets $A_1, A_2, \ldots, A_n$ where $1 \le r \le n$. This element is counted $C(r, 1)$ times by $\Sigma|A_i|$. It is counted $C(r, 2)$ times by $\Sigma|A_i \cap A_j|$. In general, it is counted $C(r, m)$ times by the summation involving m of the sets A_i. Thus, this element is counted exactly

$$C(r, 1) - C(r, 2) + C(r, 3) - \cdots + (-1)^{r+1} C(r, r)$$

times by the expression on the right-hand side of this equation. Our goal is to evaluate this quantity. By Corollary 2 of Section 6.4, we have

$$C(r, 0) - C(r, 1) + C(r, 2) - \cdots + (-1)^r C(r, r) = 0.$$

Hence,

$$1 = C(r, 0) = C(r, 1) - C(r, 2) + \cdots + (-1)^{r+1} C(r, r).$$

Therefore, each element in the union is counted exactly once by the expression on the right-hand side of the equation. This proves the principle of inclusion–exclusion. ◁

The inclusion–exclusion principle gives a formula for the number of elements in the union of n sets for every positive integer n. There are terms in this formula for the number of elements in the intersection of every nonempty subset of the collection of the n sets. Hence, there are $2^n - 1$ terms in this formula.

EXAMPLE 5 Give a formula for the number of elements in the union of four sets.

Extra Examples

Solution: The inclusion–exclusion principle shows that

$$\begin{aligned}|A_1 \cup A_2 \cup A_3 \cup A_4| = {} & |A_1| + |A_2| + |A_3| + |A_4| \\ & - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_1 \cap A_4| - |A_2 \cap A_3| - |A_2 \cap A_4| \\ & - |A_3 \cap A_4| + |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| \\ & + |A_2 \cap A_3 \cap A_4| - |A_1 \cap A_2 \cap A_3 \cap A_4|.\end{aligned}$$

Note that this formula contains 15 different terms, one for each nonempty subset of $\{A_1, A_2, A_3, A_4\}$. ◀

Exercises

1. How many elements are in $A_1 \cup A_2$ if there are 12 elements in A_1, 18 elements in A_2, and
a) $A_1 \cap A_2 = \emptyset$? **b)** $|A_1 \cap A_2| = 1$?
c) $|A_1 \cap A_2| = 6$? **d)** $A_1 \subseteq A_2$?

2. There are 345 students at a college who have taken a course in calculus, 212 who have taken a course in discrete mathematics, and 188 who have taken courses in both calculus and discrete mathematics. How many students have taken a course in either calculus or discrete mathematics?

3. A survey of households in the United States reveals that 96% have at least one television set, 98% have telephone service, and 95% have telephone service and at least one television set. What percentage of households in the United States have neither telephone service nor a television set?

4. A marketing report concerning personal computers states that 650,000 owners will buy a printer for their machines next year and 1,250,000 will buy at least one software package. If the report states that 1,450,000 owners will buy either a printer or at least one software package, how many will buy both a printer and at least one software package?

5. Find the number of elements in $A_1 \cup A_2 \cup A_3$ if there are 100 elements in each set and if
a) the sets are pairwise disjoint.
b) there are 50 common elements in each pair of sets and no elements in all three sets.
c) there are 50 common elements in each pair of sets and 25 elements in all three sets.
d) the sets are equal.

6. Find the number of elements in $A_1 \cup A_2 \cup A_3$ if there are 100 elements in A_1, 1000 in A_2, and 10,000 in A_3 if
a) $A_1 \subseteq A_2$ and $A_2 \subseteq A_3$.
b) the sets are pairwise disjoint.
c) there are two elements common to each pair of sets and one element in all three sets.

7. There are 2504 computer science students at a school. Of these, 1876 have taken a course in Java, 999 have taken a course in Linux, and 345 have taken a course in C. Further, 876 have taken courses in both Java and Linux, 231 have taken courses in both Linux and C, and 290 have taken courses in both Java and C. If 189 of these students have taken courses in Linux, Java, and C, how many of these 2504 students have not taken a course in any of these three programming languages?

8. In a survey of 270 college students, it is found that 64 like Brussels sprouts, 94 like broccoli, 58 like cauliflower, 26 like both Brussels sprouts and broccoli, 28 like both Brussels sprouts and cauliflower, 22 like both broccoli and cauliflower, and 14 like all three vegetables. How many of the 270 students do not like any of these vegetables?

9. How many students are enrolled in a course either in calculus, discrete mathematics, data structures, or programming languages at a school if there are 507, 292, 312, and 344 students in these courses, respectively; 14 in both calculus and data structures; 213 in both calculus and programming languages; 211 in both discrete mathematics and data structures; 43 in both discrete mathematics and programming languages; and no student may take calculus and discrete mathematics, or data structures and programming languages, concurrently?

10. Find the number of positive integers not exceeding 100 that are not divisible by 5 or by 7.

11. Find the number of positive integers not exceeding 1000 that are not divisible by 3, 17, or 35.

12. Find the number of positive integers not exceeding 10,000 that are not divisible by 3, 4, 7, or 11.

13. Find the number of positive integers not exceeding 100 that are either odd or the square of an integer.

14. Find the number of positive integers not exceeding 1000 that are either the square or the cube of an integer.

15. How many bit strings of length eight do not contain six consecutive 0s?

***16.** How many permutations of the 26 letters of the English alphabet do not contain any of the strings *fish*, *rat* or *bird*?

17. How many permutations of the 10 digits either begin with the 3 digits 987, contain the digits 45 in the fifth and sixth positions, or end with the 3 digits 123?

18. How many elements are in the union of four sets if each of the sets has 100 elements, each pair of the sets shares 50 elements, each three of the sets share 25 elements, and there are 5 elements in all four sets?

19. How many elements are in the union of four sets if the sets have 50, 60, 70, and 80 elements, respectively, each pair of the sets has 5 elements in common, each triple of the sets has 1 common element, and no element is in all four sets?

20. How many terms are there in the formula for the number of elements in the union of 10 sets given by the principle of inclusion–exclusion?

21. Write out the explicit formula given by the principle of inclusion–exclusion for the number of elements in the union of five sets.

22. How many elements are in the union of five sets if the sets contain 10,000 elements each, each pair of sets has 1000 common elements, each triple of sets has 100 common elements, every four of the sets have 10 common elements, and there is 1 element in all five sets?

23. Write out the explicit formula given by the principle of inclusion–exclusion for the number of elements in the union of six sets when it is known that no three of these sets have a common intersection.

***24.** Prove the principle of inclusion–exclusion using mathematical induction.

25. Let E_1, E_2, and E_3 be three events from a sample space S. Find a formula for the probability of $E_1 \cup E_2 \cup E_3$.

26. Find the probability that when a fair coin is flipped five times tails comes up exactly three times, the first and last flips come up tails, or the second and fourth flips come up heads.

27. Find the probability that when four numbers from 1 to 100, inclusive, are picked at random with no repetitions allowed, either all are odd, all are divisible by 3, or all are divisible by 5.

28. Find a formula for the probability of the union of four events in a sample space if no three of them can occur at the same time.

29. Find a formula for the probability of the union of five events in a sample space if no four of them can occur at the same time.

30. Find a formula for the probability of the union of n events in a sample space when no two of these events can occur at the same time.

31. Find a formula for the probability of the union of n events in a sample space.

8.6 Applications of Inclusion–Exclusion

8.6.1 Introduction

Many counting problems can be solved using the principle of inclusion–exclusion. For instance, we can use this principle to find the number of primes less than a positive integer. Many problems can be solved by counting the number of onto functions from one finite set to another. The inclusion–exclusion principle can be used to find the number of such functions. The well-known hatcheck problem can be solved using the principle of inclusion–exclusion. This problem asks for the probability that no person is given the correct hat back by a hatcheck person who gives the hats back randomly.

8.6.2 An Alternative Form of Inclusion–Exclusion

There is an alternative form of the principle of inclusion–exclusion that is useful in counting problems. In particular, this form can be used to solve problems that ask for the number of elements in a set that have none of n properties $P_1, P_2, \ldots, P_n$.

Let A_i be the subset containing the elements that have property P_i. The number of elements with all the properties $P_{i_1}, P_{i_2}, \ldots, P_{i_k}$ will be denoted by $N(P_{i_1} P_{i_2} \ldots P_{i_k})$. Writing these quantities in terms of sets, we have

$$|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}| = N(P_{i_1} P_{i_2} \ldots P_{i_k}).$$

If the number of elements with none of the properties $P_1, P_2, \ldots, P_n$ is denoted by $N(P_1' P_2' \ldots P_n')$ and the number of elements in the set is denoted by N, it follows that

$$N(P_1' P_2' \ldots P_n') = N - |A_1 \cup A_2 \cup \cdots \cup A_n|.$$

From the inclusion–exclusion principle, we see that

$$\begin{aligned} N(P_1' P_2' \ldots P_n') = N &- \sum_{1 \le i \le n} N(P_i) + \sum_{1 \le i < j \le n} N(P_i P_j) \\ &- \sum_{1 \le i < j < k \le n} N(P_i P_j P_k) + \cdots + (-1)^n N(P_1 P_2 \ldots P_n). \end{aligned}$$

Example 1 shows how the principle of inclusion–exclusion can be used to determine the number of solutions in integers of an equation with constraints.

EXAMPLE 1 How many solutions does

$$x_1 + x_2 + x_3 = 11$$

have, where x_1, x_2, and x_3 are nonnegative integers with $x_1 \leq 3$, $x_2 \leq 4$, and $x_3 \leq 6$?

Solution: To apply the principle of inclusion–exclusion, let a solution have property P_1 if $x_1 > 3$, property P_2 if $x_2 > 4$, and property P_3 if $x_3 > 6$. The number of solutions satisfying the inequalities $x_1 \leq 3$, $x_2 \leq 4$, and $x_3 \leq 6$ is

$$N(P_1'P_2'P_3') = N - N(P_1) - N(P_2) - N(P_3) + N(P_1P_2) + N(P_1P_3) + N(P_2P_3) - N(P_1P_2P_3).$$

Using the same techniques as in Example 5 of Section 6.5, it follows that

- N = total number of solutions $= C(3 + 11 - 1, 11) = 78$,
- $N(P_1)$ = (number of solutions with $x_1 \geq 4$) $= C(3 + 7 - 1, 7) = C(9, 7) = 36$,
- $N(P_2)$ = (number of solutions with $x_2 \geq 5$) $= C(3 + 6 - 1, 6) = C(8, 6) = 28$,
- $N(P_3)$ = (number of solutions with $x_3 \geq 7$) $= C(3 + 4 - 1, 4) = C(6, 4) = 15$,
- $N(P_1P_2)$ = (number of solutions with $x_1 \geq 4$ and $x_2 \geq 5$) $= C(3 + 2 - 1, 2) = C(4, 2) = 6$,
- $N(P_1P_3)$ = (number of solutions with $x_1 \geq 4$ and $x_3 \geq 7$) $= C(3 + 0 - 1, 0) = 1$,
- $N(P_2P_3)$ = (number of solutions with $x_2 \geq 5$ and $x_3 \geq 7$) $= 0$,
- $N(P_1P_2P_3)$ = (number of solutions with $x_1 \geq 4$, $x_2 \geq 5$, and $x_3 \geq 7$) $= 0$.

Inserting these quantities into the formula for $N(P_1'P_2'P_3')$ shows that the number of solutions with $x_1 \leq 3$, $x_2 \leq 4$, and $x_3 \leq 6$ equals

$$N(P_1'P_2'P_3') = 78 - 36 - 28 - 15 + 6 + 1 + 0 - 0 = 6.$$ ◂

8.6.3 The Sieve of Eratosthenes

In Section 4.3 we showed how to use the sieve of Eratosthenes to find all primes less than a specified positive integer n. Using the principle of inclusion–exclusion, we can find the number of primes not exceeding a specified positive integer with the same reasoning as is used in the sieve of Eratosthenes. Recall that a composite integer is divisible by a prime not exceeding its square root. So, to find the number of primes not exceeding 100, first note that composite integers not exceeding 100 must have a prime factor not exceeding 10. Because the only primes not exceeding 10 are 2, 3, 5, and 7, the primes not exceeding 100 are these four primes and those positive integers greater than 1 and not exceeding 100 that are divisible by none of 2, 3, 5, or 7. To apply the principle of inclusion–exclusion, let P_1 be the property that an integer is divisible by 2, let P_2 be the property that an integer is divisible by 3, let P_3 be the property that an integer is divisible by 5, and let P_4 be the property that an integer is divisible by 7. Thus, the number of primes not exceeding 100 is

$$4 + N(P_1'P_2'P_3'P_4').$$

Because there are 99 positive integers greater than 1 and not exceeding 100, the principle of inclusion–exclusion shows that

$$\begin{aligned}N(P_1'P_2'P_3'P_4') = 99 &- N(P_1) - N(P_2) - N(P_3) - N(P_4)\\ &+ N(P_1P_2) + N(P_1P_3) + N(P_1P_4) + N(P_2P_3) + N(P_2P_4) + N(P_3P_4)\\ &- N(P_1P_2P_3) - N(P_1P_2P_4) - N(P_1P_3P_4) - N(P_2P_3P_4)\\ &+ N(P_1P_2P_3P_4).\end{aligned}$$

The number of integers not exceeding 100 (and greater than 1) that are divisible by all the primes in a subset of $\{2, 3, 5, 7\}$ is $\lfloor 100/N \rfloor$, where N is the product of the primes in this subset. (This follows because any two of these primes have no common factor.) Consequently,

$$\begin{aligned}N(P_1'P_2'P_3'P_4') &= 99 - \left\lfloor \frac{100}{2} \right\rfloor - \left\lfloor \frac{100}{3} \right\rfloor - \left\lfloor \frac{100}{5} \right\rfloor - \left\lfloor \frac{100}{7} \right\rfloor\\ &\quad + \left\lfloor \frac{100}{2\cdot 3} \right\rfloor + \left\lfloor \frac{100}{2\cdot 5} \right\rfloor + \left\lfloor \frac{100}{2\cdot 7} \right\rfloor + \left\lfloor \frac{100}{3\cdot 5} \right\rfloor + \left\lfloor \frac{100}{3\cdot 7} \right\rfloor + \left\lfloor \frac{100}{5\cdot 7} \right\rfloor\\ &\quad - \left\lfloor \frac{100}{2\cdot 3\cdot 5} \right\rfloor - \left\lfloor \frac{100}{2\cdot 3\cdot 7} \right\rfloor - \left\lfloor \frac{100}{2\cdot 5\cdot 7} \right\rfloor - \left\lfloor \frac{100}{3\cdot 5\cdot 7} \right\rfloor + \left\lfloor \frac{100}{2\cdot 3\cdot 5\cdot 7} \right\rfloor\\ &= 99 - 50 - 33 - 20 - 14 + 16 + 10 + 7 + 6 + 4 + 2 - 3 - 2 - 1 - 0 + 0\\ &= 21.\end{aligned}$$

Hence, there are $4 + 21 = 25$ primes not exceeding 100.

8.6.4 The Number of Onto Functions

The principle of inclusion–exclusion can also be used to determine the number of onto functions from a set with m elements to a set with n elements. First consider Example 2.

EXAMPLE 2 How many onto functions are there from a set with six elements to a set with three elements?

Solution: Suppose that the elements in the codomain are b_1, b_2, and b_3. Let P_1, P_2, and P_3 be the properties that b_1, b_2, and b_3 are not in the range of the function, respectively. Note that a function is onto if and only if it has none of the properties P_1, P_2, or P_3. By the inclusion–exclusion principle it follows that the number of onto functions from a set with six elements to a set with three elements is

$$\begin{aligned}N(P_1'P_2'P_3') = N &- [N(P_1) + N(P_2) + N(P_3)]\\ &+ [N(P_1P_2) + N(P_1P_3) + N(P_2P_3)] - N(P_1P_2P_3),\end{aligned}$$

where N is the total number of functions from a set with six elements to one with three elements. We will evaluate each of the terms on the right-hand side of this equation.

From Example 6 of Section 6.1, it follows that $N = 3^6$. Note that $N(P_i)$ is the number of functions that do not have b_i in their range. Hence, there are two choices for the value of the function at each element of the domain. Therefore, $N(P_i) = 2^6$. Furthermore, there are $C(3, 1)$ terms of this kind. Note that $N(P_iP_j)$ is the number of functions that do not have b_i and b_j in their range. Hence, there is only one choice for the value of the function at each element of the domain. Therefore, $N(P_iP_j) = 1^6 = 1$. Furthermore, there are $C(3, 2)$ terms of this kind. Also, note that $N(P_1P_2P_3) = 0$, because this term is the number of functions that have none of b_1, b_2, and b_3 in their range. Clearly, there are no such functions, so the number of onto functions from a set with six elements to one with three elements is

$$3^6 - C(3, 1)2^6 + C(3, 2)1^6 = 729 - 192 + 3 = 540.$$ ◀

The general result that tells us how many onto functions there are from a set with m elements to one with n elements will now be stated. The proof of this result is left as an exercise for the reader.

THEOREM 1 Let m and n be positive integers with $m \geq n$. Then, there are

$$n^m - C(n, 1)(n-1)^m + C(n, 2)(n-2)^m - \cdots + (-1)^{n-1}C(n, n-1) \cdot 1^m$$

onto functions from a set with m elements to a set with n elements.

Counting onto functions is much harder than counting one-to-one functions!

An onto function from a set with m elements to a set with n elements corresponds to a way to distribute the m elements in the domain to n indistinguishable boxes so that no box is empty, and then to associate each of the n elements of the codomain to a box. This means that the number of onto functions from a set with m elements to a set with n elements is the number of ways to distribute m distinguishable objects to n indistinguishable boxes so that no box is empty multiplied by the number of permutations of a set with n elements. Consequently, the number of onto functions from a set with m elements to a set with n elements equals $n!S(m, n)$, where $S(m, n)$ is a *Stirling number of the second kind* defined in Section 6.5. This means that we can use Theorem 1 to deduce the formula given in Section 6.5 for $S(m, n)$. (See Chapter 6 of [MiRo91] for more details about Stirling numbers of the second kind.)

One of the many different applications of Theorem 1 will now be described.

EXAMPLE 3 How many ways are there to assign five different jobs to four different employees if every employee is assigned at least one job?

Solution: Consider the assignment of jobs as a function from the set of five jobs to the set of four employees. An assignment where every employee gets at least one job is the same as an onto function from the set of jobs to the set of employees. Hence, by Theorem 1 it follows that there are

$$4^5 - C(4, 1)3^5 + C(4, 2)2^5 - C(4, 3)1^5 = 1024 - 972 + 192 - 4 = 240$$

ways to assign the jobs so that each employee is assigned at least one job. ◀

8.6.5 Derangements

The principle of inclusion–exclusion will be used to count the permutations of n objects that leave no objects in their original positions. Consider Example 4.

EXAMPLE 4 **The Hatcheck Problem** A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the probability that no one receives the correct hat? ◀

Remark: The answer is the number of ways the hats can be arranged so that there is no hat in its original position divided by $n!$, the number of permutations of n hats. We will return to this example after we find the number of permutations of n objects that leave no objects in their original position.

Links

A **derangement** is a permutation of objects that leaves no object in its original position. To solve the problem posed in Example 4 we will need to determine the number of derangements of a set of n objects.

EXAMPLE 5 The permutation 21453 is a derangement of 12345 because no number is left in its original position. However, 21543 is not a derangement of 12345, because this permutation leaves 4 fixed. ◀

Let D_n denote the number of derangements of n objects. For instance, $D_3 = 2$, because the derangements of 123 are 231 and 312. We will evaluate D_n, for all positive integers n, using the principle of inclusion–exclusion.

THEOREM 2 The number of derangements of a set with n elements is

$$D_n = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right].$$

Proof: Let a permutation have property P_i if it fixes element i. The number of derangements is the number of permutations having none of the properties P_i for $i = 1, 2, \ldots, n$. This means that

$$D_n = N(P_1'P_2' \ldots P_n').$$

Using the principle of inclusion–exclusion, it follows that

$$D_n = N - \sum_i N(P_i) + \sum_{i<j} N(P_iP_j) - \sum_{i<j<k} N(P_iP_jP_k) + \cdots + (-1)^n N(P_1P_2 \ldots P_n),$$

where N is the number of permutations of n elements. This equation states that the number of permutations that fix no elements equals the total number of permutations, less the number that fix at least one element, plus the number that fix at least two elements, less the number that fix at least three elements, and so on. All the quantities that occur on the right-hand side of this equation will now be found.

First, note that $N = n!$, because N is simply the total number of permutations of n elements. Also, $N(P_i) = (n-1)!$. This follows from the product rule, because $N(P_i)$ is the number of permutations that fix element i, so the ith position of the permutation is determined, but each of the remaining positions can be filled arbitrarily. Similarly,

$$N(P_iP_j) = (n-2)!,$$

Links

HISTORICAL NOTE In *rencontres* (matches), an old French card game, the 52 cards in a deck are laid out in a row. The cards of a second deck are laid out with one card of the second deck on top of each card of the first deck. The score is determined by counting the number of matching cards in the two decks. In 1708 Pierre Raymond de Montmort (1678–1719) posed *le problème de rencontres:* What is the probability that no matches take place in the game of rencontres? The solution to Montmort's problem is the probability that a randomly selected permutation of 52 objects is a derangement, namely, $D_{52}/52!$, which, as we will see, is approximately $1/e$.

because this is the number of permutations that fix elements i and j, but where the other $n-2$ elements can be arranged arbitrarily. In general, note that

$$N(P_{i_1}P_{i_2}\ldots P_{i_m}) = (n-m)!,$$

because this is the number of permutations that fix elements $i_1, i_2, \ldots, i_m$, but where the other $n-m$ elements can be arranged arbitrarily. Because there are $C(n, m)$ ways to choose m elements from n, it follows that

$$\sum_{1\le i\le n} N(P_i) = C(n, 1)(n-1)!,$$

$$\sum_{1\le i<j\le n} N(P_iP_j) = C(n, 2)(n-2)!,$$

and in general,

$$\sum_{1\le i_1<i_2<\cdots<i_m\le n} N(P_{i_1}P_{i_2}\ldots P_{i_m}) = C(n, m)(n-m)!.$$

Consequently, inserting these quantities into our formula for D_n gives

$$\begin{aligned}D_n &= n! - C(n, 1)(n-1)! + C(n, 2)(n-2)! - \cdots + (-1)^n C(n, n)(n-n)!\\ &= n! - \frac{n!}{1!(n-1)!}(n-1)! + \frac{n!}{2!(n-2)!}(n-2)! - \cdots + (-1)^n \frac{n!}{n!\,0!}0!.\end{aligned}$$

Simplifying this expression gives

$$D_n = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n\frac{1}{n!}\right].$$

◁

It is now straightforward to find D_n for a given positive integer n. For instance, using Theorem 2, it follows that

$$D_3 = 3!\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!}\right] = 6\left(1 - 1 + \frac{1}{2} - \frac{1}{6}\right) = 2,$$

as we have previously remarked.

The solution of the problem in Example 4 can now be given.

Solution: The probability that no one receives the correct hat is $D_n/n!$. By Theorem 2, this probability is

$$\frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n\frac{1}{n!}.$$

The values of this probability for $2 \le n \le 7$ are displayed in Table 1.

TABLE 1 The Probability of a Derangement.

n	2	3	4	5	6	7
$D_n/n!$	0.50000	0.33333	0.37500	0.36667	0.36806	0.36786

By the identity $e^x = \sum_{j=0}^{\infty} x^j/j!$ for all real numbers x (from calculus), we know that

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n \frac{1}{n!} + \cdots \approx 0.368.$$

Because this is an alternating series with terms tending to zero, it follows that as n grows without bound, the probability that no one receives the correct hat converges to $e^{-1} \approx 0.368$. In fact, this probability can be shown to be within $1/(n+1)!$ of e^{-1}. ◀

Exercises

1. Suppose that in a bushel of 100 apples there are 20 that have worms in them and 15 that have bruises. Only those apples with neither worms nor bruises can be sold. If there are 10 bruised apples that have worms in them, how many of the 100 apples can be sold?

2. Of 1000 applicants for a mountain-climbing trip in the Himalayas, 450 get altitude sickness, 622 are not in good enough shape, and 30 have allergies. An applicant qualifies if and only if this applicant does not get altitude sickness, is in good shape, and does not have allergies. If there are 111 applicants who get altitude sickness and are not in good enough shape, 14 who get altitude sickness and have allergies, 18 who are not in good enough shape and have allergies, and 9 who get altitude sickness, are not in good enough shape, and have allergies, how many applicants qualify?

3. How many solutions does the equation $x_1 + x_2 + x_3 = 13$ have where x_1, x_2, and x_3 are nonnegative integers less than 6?

4. Find the number of solutions of the equation $x_1 + x_2 + x_3 + x_4 = 17$, where x_i, $i = 1, 2, 3, 4$, are nonnegative integers such that $x_1 \leq 3$, $x_2 \leq 4$, $x_3 \leq 5$, and $x_4 \leq 8$.

5. Find the number of primes less than 200 using the principle of inclusion–exclusion.

6. An integer is called **squarefree** if it is not divisible by the square of a positive integer greater than 1. Find the number of squarefree positive integers less than 100.

7. How many positive integers less than 10,000 are not the second or higher power of an integer?

8. How many onto functions are there from a set with seven elements to one with five elements?

9. How many ways are there to distribute six different toys to three different children such that each child gets at least one toy?

10. In how many ways can eight distinct balls be distributed into three distinct urns if each urn must contain at least one ball?

11. In how many ways can seven different jobs be assigned to four different employees so that each employee is assigned at least one job and the most difficult job is assigned to the best employee?

12. List all the derangements of $\{1, 2, 3, 4\}$.

13. How many derangements are there of a set with seven elements?

14. What is the probability that none of 10 people receives the correct hat if a hatcheck person hands their hats back randomly?

15. A machine that inserts letters into envelopes goes haywire and inserts letters randomly into envelopes. What is the probability that in a group of 100 letters

a) no letter is put into the correct envelope?
b) exactly one letter is put into the correct envelope?
c) exactly 98 letters are put into the correct envelopes?
d) exactly 99 letters are put into the correct envelopes?
e) all letters are put into the correct envelopes?

16. A group of n students is assigned seats for each of two classes in the same classroom. How many ways can these seats be assigned if no student is assigned the same seat for both classes?

***17.** How many ways can the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 be arranged so that no even digit is in its original position?

***18.** Use a combinatorial argument to show that the sequence $\{D_n\}$, where D_n denotes the number of derangements of n objects, satisfies the recurrence relation

$$D_n = (n-1)(D_{n-1} + D_{n-2})$$

for $n \geq 2$. [*Hint:* Note that there are $n - 1$ choices for the first element k of a derangement. Consider separately the derangements that start with k that do and do not have 1 in the kth position.]

***19.** Use Exercise 18 to show that

$$D_n = nD_{n-1} + (-1)^n$$

for $n \geq 1$.

20. Use Exercise 19 to find an explicit formula for D_n.

21. For which positive integers n is D_n, the number of derangements of n objects, even?

22. Suppose that p and q are distinct primes. Use the principle of inclusion–exclusion to find $\phi(pq)$, the number of positive integers not exceeding pq that are relatively prime to pq.

***23.** Use the principle of inclusion–exclusion to derive a formula for $\phi(n)$ when the prime factorization of n is

$$n = p_1^{a_1} p_2^{a_2} \cdots p_m^{a_m}.$$

*24. Show that if n is a positive integer, then

$$n! = C(n, 0)D_n + C(n, 1)D_{n-1} + \cdots + C(n, n-1)D_1 + C(n, n)D_0,$$

where D_k is the number of derangements of k objects.

25. How many derangements of $\{1, 2, 3, 4, 5, 6\}$ begin with the integers 1, 2, and 3, in some order?

26. How many derangements of $\{1, 2, 3, 4, 5, 6\}$ end with the integers 1, 2, and 3, in some order?

27. Prove Theorem 1.

Key Terms and Results

TERMS

recurrence relation: a formula expressing terms of a sequence, except for some initial terms, as a function of one or more previous terms of the sequence

initial conditions for a recurrence relation: the values of the terms of a sequence satisfying the recurrence relation before this relation takes effect

dynamic programming: an algorithmic paradigm that finds the solution to an optimization problem by recursively breaking down the problem into overlapping subproblems and combining their solutions with the help of a recurrence relation

linear homogeneous recurrence relation with constant coefficients: a recurrence relation that expresses the terms of a sequence, except initial terms, as a linear combination of previous terms

characteristic roots of a linear homogeneous recurrence relation with constant coefficients: the roots of the polynomial associated with a linear homogeneous recurrence relation with constant coefficients

linear nonhomogeneous recurrence relation with constant coefficients: a recurrence relation that expresses the terms of a sequence, except for initial terms, as a linear combination of previous terms plus a function that is not identically zero that depends only on the index

divide-and-conquer algorithm: an algorithm that solves a problem recursively by splitting it into a fixed number of smaller nonoverlapping subproblems of the same type

generating function of a sequence: the formal series that has the nth term of the sequence as the coefficient of x^n

sieve of Eratosthenes: a procedure for finding the primes less than a specified positive integer

derangement: a permutation of objects such that no object is in its original place

RESULTS

the formula for the number of elements in the union of two finite sets:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

the formula for the number of elements in the union of three finite sets:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

the principle of inclusion–exclusion:

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{1 \le i \le n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1} |A_1 \cap A_2 \cap \cdots \cap A_n|$$

the number of onto functions from a set with m elements to a set with n elements:

$$n^m - C(n, 1)(n-1)^m + C(n, 2)(n-2)^m - \cdots + (-1)^{n-1} C(n, n-1) \cdot 1^m$$

the number of derangements of n objects:

$$D_n = n! \left[1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n \frac{1}{n!}\right]$$

Review Questions

1. a) What is a recurrence relation?
   b) Find a recurrence relation for the amount of money that will be in an account after n years if \$1,000,000 is deposited in an account yielding 9% annually.

2. Explain how the Fibonacci numbers are used to solve Fibonacci's problem about rabbits.

3. a) Find a recurrence relation for the number of steps needed to solve the Tower of Hanoi puzzle.
   b) Show how this recurrence relation can be solved using iteration.

4. a) Explain how to find a recurrence relation for the number of bit strings of length n not containing two consecutive 1s.
   b) Describe another counting problem that has a solution satisfying the same recurrence relation.

5. a) What is dynamic programming and how are recurrence relations used in algorithms that follow this paradigm?

b) Explain how dynamic programming can be used to schedule talks in a lecture hall from a set of possible talks to maximize overall attendance.

6. Define a linear homogeneous recurrence relation of degree k.

7. a) Explain how to solve linear homogeneous recurrence relations of degree 2.

b) Solve the recurrence relation $a_n = 13a_{n-1} - 22a_{n-2}$ for $n \geq 2$ if $a_0 = 3$ and $a_1 = 15$.

c) Solve the recurrence relation $a_n = 14a_{n-1} - 49a_{n-2}$ for $n \geq 2$ if $a_0 = 3$ and $a_1 = 35$.

8. a) Explain how to find $f(b^k)$ where k is a positive integer if $f(n)$ satisfies the divide-and-conquer recurrence relation $f(n) = af(n/b) + g(n)$ whenever b divides the positive integer n.

b) Find $f(256)$ if $f(n) = 3f(n/4) + 5n/4$ and $f(1) = 7$.

9. a) Derive a divide-and-conquer recurrence relation for the number of comparisons used to find a number in a list using a binary search.

b) Give a big-O estimate for the number of comparisons used by a binary search from the divide-and-conquer recurrence relation you gave in (a) using Theorem 1 in Section 8.3.

10. a) Give a formula for the number of elements in the union of three sets.

b) Explain why this formula is valid.

c) Explain how to use the formula from (a) to find the number of integers not exceeding 1000 that are divisible by 6, 10, or 15.

d) Explain how to use the formula from (a) to find the number of solutions in nonnegative integers to the equation $x_1 + x_2 + x_3 + x_4 = 22$ with $x_1 < 8$, $x_2 < 6$, and $x_3 < 5$.

11. a) Give a formula for the number of elements in the union of four sets and explain why it is valid.

b) Suppose the sets A_1, A_2, A_3, and A_4 each contain 25 elements, the intersection of any two of these sets contains 5 elements, the intersection of any three of these sets contains 2 elements, and 1 element is in all four of the sets. How many elements are in the union of the four sets?

12. a) State the principle of inclusion–exclusion.

b) Outline a proof of this principle.

13. Explain how the principle of inclusion–exclusion can be used to count the number of onto functions from a set with m elements to a set with n elements.

14. a) How can you count the number of ways to assign m jobs to n employees so that each employee is assigned at least one job?

b) How many ways are there to assign seven jobs to three employees so that each employee is assigned at least one job?

15. Explain how the inclusion–exclusion principle can be used to count the number of primes not exceeding the positive integer n.

16. a) Define a derangement.

b) Why is counting the number of ways a hatcheck person can return hats to n people, so that no one receives the correct hat, the same as counting the number of derangements of n objects?

c) Explain how to count the number of derangements of n objects.

Supplementary Exercises

1. A group of 10 people begin a chain letter, with each person sending the letter to four other people. Each of these people sends the letter to four additional people.

a) Find a recurrence relation for the number of letters sent at the nth stage of this chain letter, if no person ever receives more than one letter.

b) What are the initial conditions for the recurrence relation in part (a)?

c) How many letters are sent at the nth stage of the chain letter?

2. A nuclear reactor has created 18 grams of a particular radioactive isotope. Every hour 1% of this radioactive isotope decays.

a) Set up a recurrence relation for the amount of this isotope left n hours after its creation.

b) What are the initial conditions for the recurrence relation in part (a)?

c) Solve this recurrence relation.

3. Every hour the U.S. government prints 10,000 more \$1 bills, 4000 more \$5 bills, 3000 more \$10 bills, 2500 more \$20 bills, 1000 more \$50 bills, and the same number of \$100 bills as it did the previous hour. In the initial hour 1000 of each bill were produced.

a) Set up a recurrence relation for the amount of money produced in the nth hour.

b) What are the initial conditions for the recurrence relation in part (a)?

c) Solve the recurrence relation for the amount of money produced in the nth hour.

d) Set up a recurrence relation for the total amount of money produced in the first n hours.

e) Solve the recurrence relation for the total amount of money produced in the first n hours.

4. Suppose that every hour there are two new bacteria in a colony for each bacterium that was present the previous hour, and that all bacteria 2 hours old die. The colony starts with 100 new bacteria.

a) Set up a recurrence relation for the number of bacteria present after n hours.

b) What is the solution of this recurrence relation?

c) When will the colony contain more than 1 million bacteria?

5. Messages are sent over a communications channel using two different signals. One signal requires 2 microseconds for transmittal, and the other signal requires 3 microseconds for transmittal. Each signal of a message is followed immediately by the next signal.

a) Find a recurrence relation for the number of different signals that can be sent in n microseconds.

b) What are the initial conditions of the recurrence relation in part (a)?

c) How many different messages can be sent in 12 microseconds?

6. A small post office has only 4-cent stamps, 6-cent stamps, and 10-cent stamps. Find a recurrence relation for the number of ways to form postage of n cents with these stamps if the order that the stamps are used matters. What are the initial conditions for this recurrence relation?

7. How many ways are there to form these postages using the rules described in Exercise 6?

a) 12 cents **b)** 14 cents

c) 18 cents **d)** 22 cents

8. Find the solutions of the simultaneous system of recurrence relations

$$a_n = a_{n-1} + b_{n-1}$$
$$b_n = a_{n-1} - b_{n-1}$$

with $a_0 = 1$ and $b_0 = 2$.

9. Solve the recurrence relation $a_n = a_{n-1}^2/a_{n-2}$ if $a_0 = 1$ and $a_1 = 2$. [*Hint:* Take logarithms of both sides to obtain a recurrence relation for the sequence $\log a_n$, $n = 0, 1, 2, \ldots$.]

***10.** Solve the recurrence relation $a_n = a_{n-1}^3 a_{n-2}^2$ if $a_0 = 2$ and $a_1 = 2$. (See the hint for Exercise 9.)

11. Find the solution of the recurrence relation $a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3} + 1$ if $a_0 = 2$, $a_1 = 4$, and $a_2 = 8$.

12. Find the solution of the recurrence relation $a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3}$ if $a_0 = 2$, $a_1 = 2$, and $a_2 = 4$.

***13.** Suppose that in Example 1 of Section 8.1 a pair of rabbits leaves the island after reproducing twice. Find a recurrence relation for the number of rabbits on the island in the middle of the nth month.

***14.** In this exercise we construct a dynamic programming algorithm for solving the problem of finding a subset S of items chosen from a set of n items where item i has a weight w_i, which is a positive integer, so that the total weight of the items in S is a maximum but does not exceed a fixed weight limit W. Let $M(j, w)$ denote the maximum total weight of the items in a subset of the first j items such that this total weight does not exceed w. This problem is known as the **knapsack problem**.

a) Show that if $w_j > w$, then $M(j, w) = M(j-1, w)$.

b) Show that if $w_j \le w$, then $M(j, w) = \max(M(j-1, w), w_j + M(j-1, w-w_j))$.

c) Use (a) and (b) to construct a dynamic programming algorithm for determining the maximum total weight of items so that this total weight does not exceed W. In your algorithm store the values $M(j, w)$ as they are found.

d) Explain how you can use the values $M(j, w)$ computed by the algorithm in part (c) to find a subset of items with maximum total weight not exceeding W.

In Exercises 15–18 we develop a dynamic programming algorithm for finding a longest common subsequence of two sequences $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$, an important problem in the comparison of DNA of different organisms.

15. Suppose that $c_1, c_2, \ldots, c_p$ is a longest common subsequence of the sequences $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$.

a) Show that if $a_m = b_n$, then $c_p = a_m = b_n$ and $c_1, c_2, \ldots, c_{p-1}$ is a longest common subsequence of $a_1, a_2, \ldots, a_{m-1}$ and $b_1, b_2, \ldots, b_{n-1}$ when $p > 1$.

b) Suppose that $a_m \ne b_n$. Show that if $c_p \ne a_m$, then $c_1, c_2, \ldots, c_p$ is a longest common subsequence of $a_1, a_2, \ldots, a_{m-1}$ and $b_1, b_2, \ldots, b_n$ and also show that if $c_p \ne b_n$, then $c_1, c_2, \ldots, c_p$ is a longest common subsequence of $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_{n-1}$.

16. Let $L(i, j)$ denote the length of a longest common subsequence of $a_1, a_2, \ldots, a_i$ and $b_1, b_2, \ldots, b_j$, where $0 \le i \le m$ and $0 \le j \le n$. Use parts (a) and (b) of Exercise 15 to show that $L(i, j)$ satisfies the recurrence relation $L(i, j) = L(i-1, j-1) + 1$ if both i and j are nonzero and $a_i = b_i$, and $L(i, j) = \max(L(i, j-1), L(i-1, j))$ if both i and j are nonzero and $a_i \ne b_i$, and the initial condition $L(i, j) = 0$ if $i = 0$ or $j = 0$.

17. Use Exercise 16 to construct a dynamic programming algorithm for computing the length of a longest common subsequence of two sequences $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$, storing the values of $L(i, j)$ as they are found.

18. Develop an algorithm for finding a longest common subsequence of two sequences $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$ using the values $L(i, j)$ found by the algorithm in Exercise 17.

19. Find the solution to the recurrence relation $f(n) = f(n/2) + n^2$ for $n = 2^k$ where k is a positive integer and $f(1) = 1$.

20. Find the solution to the recurrence relation $f(n) = 3f(n/5) + 2n^4$, when n is divisible by 5, for $n = 5^k$, where k is a positive integer and $f(1) = 1$.

21. Give a big-O estimate for the size of f in Exercise 20 if f is an increasing function.

22. Find a recurrence relation that describes the number of comparisons used by the following algorithm: Find the largest and second largest elements of a sequence of n numbers recursively by splitting the sequence into two subsequences with an equal number of terms, or where there is one more term in one subsequence than in the other, at each stage. Stop when subsequences with two terms are reached.

23. Give a big-O estimate for the number of comparisons used by the algorithm described in Exercise 22.

24. A sequence $a_1, a_2, \ldots, a_n$ is **unimodal** if and only if there is an index m, $1 \le m \le n$, such that $a_i < a_{i+1}$ when $1 \le i < m$ and $a_i > a_{i+1}$ when $m \le i < n$. That is, the terms of the sequence strictly increase until the mth term and they strictly decrease after it, which implies that a_m is the largest term. In this exercise, a_m will always denote the largest term of the unimodal sequence $a_1, a_2, \ldots, a_n$.

a) Show that a_m is the unique term of the sequence that is greater than both the term immediately preceding it and the term immediately following it.

b) Show that if $a_i < a_{i+1}$ where $1 \le i < n$, then $i + 1 \le m \le n$.

c) Show that if $a_i > a_{i+1}$ where $1 \le i < n$, then $1 \le m \le i$.

d) Develop a divide-and-conquer algorithm for locating the index m. [*Hint:* Suppose that $i < m < j$. Use parts (a), (b), and (c) to determine whether $\lfloor (i+j)/2 \rfloor + 1 \le m \le n$, $1 \le m \le \lfloor (i+j)/2 \rfloor - 1$, or $m = \lfloor (i+j)/2 \rfloor$.]

25. Show that the algorithm from Exercise 24 has worst-case time complexity $O(\log n)$ in terms of the number of comparisons.

Let $\{a_n\}$ be a sequence of real numbers. The **forward differences** of this sequence are defined recursively as follows: The **first forward difference** is $\Delta a_n = a_{n+1} - a_n$; the **$(k+1)$st forward difference** $\Delta^{k+1} a_n$ is obtained from $\Delta^k a_n$ by $\Delta^{k+1} a_n = \Delta^k a_{n+1} - \Delta^k a_n$.

26. Find Δa_n, where

a) $a_n = 3$. **b)** $a_n = 4n + 7$. **c)** $a_n = n^2 + n + 1$.

27. Let $a_n = 3n^3 + n + 2$. Find $\Delta^k a_n$, where k equals

a) 2. **b)** 3. **c)** 4.

***28.** Suppose that $a_n = P(n)$, where P is a polynomial of degree d. Prove that $\Delta^{d+1} a_n = 0$ for all nonnegative integers n.

29. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. Show that

$$\Delta(a_n b_n) = a_{n+1}(\Delta b_n) + b_n(\Delta a_n).$$

30. Show that if $F(x)$ and $G(x)$ are the generating functions for the sequences $\{a_k\}$ and $\{b_k\}$, respectively, and c and d are real numbers, then $(cF + dG)(x)$ is the generating function for $\{ca_k + db_k\}$.

31. (*Requires calculus*) This exercise shows how generating functions can be used to solve the recurrence relation $(n + 1)a_{n+1} = a_n + (1/n!)$ for $n \ge 0$ with initial condition $a_0 = 1$.

a) Let $G(x)$ be the generating function for $\{a_n\}$. Show that $G'(x) = G(x) + e^x$ and $G(0) = 1$.

b) Show from part (a) that $(e^{-x} G(x))' = 1$, and conclude that $G(x) = xe^x + e^x$.

c) Use part (b) to find a closed form for a_n.

32. Suppose that 14 students receive an A on the first exam in a discrete mathematics class, and 18 receive an A on the second exam. If 22 students received an A on either the first exam or the second exam, how many students received an A on both exams?

33. There are 323 farms in Monmouth County that have at least one of horses, cows, and sheep. If 224 have horses, 85 have cows, 57 have sheep, and 18 farms have all three types of animals, how many farms have exactly two of these three types of animals?

34. Queries to a database of student records at a college produced the following data: There are 2175 students at the college, 1675 of these are not freshmen, 1074 students have taken a course in calculus, 444 students have taken a course in discrete mathematics, 607 students are not freshmen and have taken calculus, 350 students have taken calculus and discrete mathematics, 201 students are not freshmen and have taken discrete mathematics, and 143 students are not freshmen and have taken both calculus and discrete mathematics. Can all the responses to the queries be correct?

35. Students in the school of mathematics at a university major in one or more of the following four areas: applied mathematics (AM), pure mathematics (PM), operations research (OR), and computer science (CS). How many students are in this school if (including joint majors) there are 23 students majoring in AM; 17 in PM; 44 in OR; 63 in CS; 5 in AM and PM; 8 in AM and CS; 4 in AM and OR; 6 in PM and CS; 5 in PM and OR; 14 in OR and CS; 2 in PM, OR, and CS; 2 in AM, OR, and CS; 1 in PM, AM, and OR; 1 in PM, AM, and CS; and 1 in all four fields.

36. How many terms are needed when the inclusion–exclusion principle is used to express the number of elements in the union of seven sets if no more than five of these sets have a common element?

37. How many solutions in positive integers are there to the equation $x_1 + x_2 + x_3 = 20$ with $2 < x_1 < 6$, $6 < x_2 < 10$, and $0 < x_3 < 5$?

38. How many positive integers less than 1,000,000 are

a) divisible by 2, 3, or 5?

b) not divisible by 7, 11, or 13?

c) divisible by 3 but not by 7?

39. How many positive integers less than 200 are

a) second or higher powers of integers?

b) either primes or second or higher powers of integers?

c) not divisible by the square of an integer greater than 1?

d) not divisible by the cube of an integer greater than 1?

e) not divisible by three or more primes?

*40. How many ways are there to assign six different jobs to three different employees if the hardest job is assigned to the most experienced employee and the easiest job is assigned to the least experienced employee?

41. What is the probability that exactly one person is given back the correct hat by a hatcheck person who gives n people their hats back at random?

42. How many bit strings of length six do not contain four consecutive 1s?

43. What is the probability that a bit string of length six chosen at random contains at least four 1s?

Computer Projects

Write programs with these input and output.

1. Given a positive integer n, list all the moves required in the Tower of Hanoi puzzle to move n disks from one peg to another according to the rules of the puzzle.
2. Given a positive integer n and an integer k with $1 \leq k \leq n$, list all the moves used by the Frame–Stewart algorithm (described in the preamble to Exercise 38 of Section 8.1) to move n disks from one peg to another using four pegs according to the rules of the puzzle.
3. Given a positive integer n, list all the bit sequences of length n that do not have a pair of consecutive 0s.
4. Given an integer n greater than 1, write out all ways to parenthesize the product of $n + 1$ variables.
5. Given a set of n talks, their start and end times, and the number of attendees at each talk, use dynamic programming to schedule a subset of these talks in a single lecture hall to maximize total attendance.
6. Given matrices $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A}_n$, with dimensions $m_1 \times m_2, m_2 \times m_3, \ldots, m_n \times m_{n+1}$, respectively, each with integer entries, use dynamic programming, as outlined in Exercise 57 in Section 8.1, to find the minimum number of multiplications of integers needed to compute $\mathbf{A_1 A_2} \cdots \mathbf{A}_n$.
7. Given a recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$, where c_1 and c_2 are real numbers, initial conditions $a_0 = C_0$ and $a_1 = C_1$, and a positive integer k, find a_k using iteration.
8. Given a recurrence relation $a_n = c_1 a_{n-1} + c_2 a_{n-2}$ and initial conditions $a_0 = C_0$ and $a_1 = C_1$, determine the unique solution.
9. Given a recurrence relation of the form $f(n) = af(n/b) + c$, where a is a real number, b is a positive integer, and c is a real number, and a positive integer k, find $f(b^k)$ using iteration.
10. Given the number of elements in the intersection of three sets, the number of elements in each pairwise intersection of these sets, and the number of elements in each set, find the number of elements in their union.
11. Given a positive integer n, produce the formula for the number of elements in the union of n sets.
12. Given positive integers m and n, find the number of onto functions from a set with m elements to a set with n elements.
13. Given a positive integer n, list all the derangements of the set $\{1, 2, 3, \ldots, n\}$.

Computations and Explorations

Use a computational program or programs you have written to do these exercises.

1. Find the exact value of f_{100}, f_{500}, and f_{1000}, where f_n is the nth Fibonacci number.
2. Find the smallest Fibonacci number greater than 1,000,000, greater than 1,000,000,000, and greater than 1,000,000,000,000.
3. Find as many prime Fibonacci numbers as you can. It is unknown whether there are infinitely many of these.
4. Write out all the moves required to solve the Tower of Hanoi puzzle with 10 disks.
5. Write out all the moves required to use the Frame–Stewart algorithm to move 20 disks from one peg to another peg using four pegs according to the rules of the Reve's puzzle.
6. Verify the Frame conjecture for solving the Reve's puzzle for n disks for as many integers n as possible by showing that the puzzle cannot be solved using fewer moves than are made by the Frame–Stewart algorithm with the optimal choice of k.
7. Compute the number of operations required to multiply two integers with n bits for various integers n including 16, 64, 256, and 1024 using the fast multiplication described in Example 4 of Section 8.3 and the standard algorithm for multiplying integers (Algorithm 3 in Section 4.2).
8. Compute the number of operations required to multiply two $n \times n$ matrices for various integers n including 4, 16, 64, and 128 using the fast matrix multiplication described

in Example 5 of Section 8.3 and the standard algorithm for multiplying matrices (Algorithm 1 in Section 3.3).

9. Find the number of primes not exceeding 10,000 using the method described in Section 8.6 to find the number of primes not exceeding 100.

10. List all the derangements of $\{1, 2, 3, 4, 5, 6, 7, 8\}$.

11. Compute the probability that a permutation of n objects is a derangement for all positive integers not exceeding 20 and determine how quickly these probabilities approach the number $1/e$.

Writing Projects

Respond to these with essays using outside sources.

1. Find the original source where Fibonacci presented his puzzle about modeling rabbit populations. Discuss this problem and other problems posed by Fibonacci and give some information about Fibonacci himself.

2. Explain how the Fibonacci numbers arise in a variety of applications, such as in phyllotaxis, the study of arrangement of leaves in plants, in the study of reflections by mirrors, and so on.

3. Describe different variations of the Tower of Hanoi puzzle, including those with more than three pegs (including the Reve's puzzle discussed in the text and exercises), those where disk moves are restricted, and those where disks may have the same size. Include what is known about the number of moves required to solve each variation.

4. Discuss as many different problems as possible where the Catalan numbers arise.

5. Discuss some of the problems in which Richard Bellman first used dynamic programming.

6. Describe the role dynamic programming algorithms play in bioinformatics including for DNA sequence comparison, gene comparison, and RNA structure prediction.

7. Describe the use of dynamic programming in economics including its use to study optimal consumption and saving.

8. Explain how dynamic programming can be used to solve the egg-dropping puzzle which determines which floors of a multistory building it is safe to drop eggs from without breaking.

9. Describe the solution of Ulam's problem (see Exercise 28 in Section 8.3) involving searching with one lie found by Andrzej Pelc.

10. Discuss variations of Ulam's problem (see Exercise 28 in Section 8.3) involving searching with more than one lie and what is known about this problem.

11. Define the convex hull of a set of points in the plane and describe three different algorithms, including a divide-and-conquer algorithm, for finding the convex hull of a set of points in the plane.

12. Describe how sieve methods are used in number theory. What kind of results have been established using such methods?

13. Look up the rules of the old French card game of *rencontres*. Describe these rules and describe the work of Pierre Raymond de Montmort on *le problème de rencontres*.

14. Describe how exponential generating functions can be used to solve a variety of counting problems.

15. Describe the Polyá theory of counting and the kind of counting problems that can be solved using this theory.

16. The *problème des ménages* (the problem of the households) asks for the number of ways to arrange n couples around a table so that the sexes alternate and no husband and wife are seated together. Explain the method used by E. Lucas to solve this problem.

17. Explain how *rook polynomials* can be used to solve counting problems.