1

The Mathematics of Choice

> It seems that mathematical ideas are arranged somehow in strata, the ideas in each stratum being linked by a complex of relations both among themselves and with those above and below. The lower the stratum, the deeper (and in general the more difficult) the idea. Thus, the idea of an irrational is deeper than the idea of an integer.
>
> — G. H. Hardy (*A Mathematician's Apology*)

Roughly speaking, the first chapter of this book is the top stratum, the surface layer of combinatorics. Even so, it is far from superficial. While the first main result, the so-called fundamental counting principle, is nearly self-evident, it has enormous implications throughout combinatorial enumeration. In the version presented here, one is faced with a sequence of decisions, each of which involves some number of choices. It is from situations like this that the chapter derives its name.

To the uninitiated, mathematics may appear to be "just so many numbers and formulas." In fact, the numbers and formulas should be regarded as shorthand notes, summarizing *ideas*. Some ideas from the first section are summarized by an algebraic formula for multinomial coefficients. Special cases of these numbers are addressed from a combinatorial perspective in Section 1.2.

Section 1.3 is an optional discussion of probability theory which can be omitted if probabilistic exercises in subsequent sections are approached with caution. Section 1.4 is an optional excursion into the theory of binary codes which can be omitted by those not planning to visit Chapter 6. Sections 1.3 and 1.4 are partly motivational, illustrating that even the most basic combinatorial ideas have real-life applications.

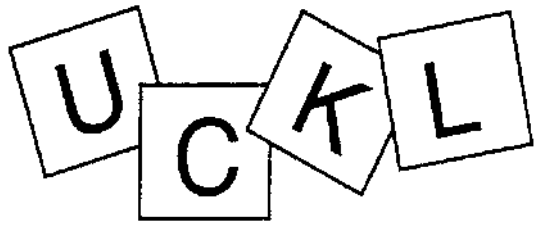
In Section 1.5, ideas behind the formulas for sums of powers of positive integers motivate the study of relations among binomial coefficients. Choice is again the topic in Section 1.6, this time with or without replacement, where order does or doesn't matter.

To better organize and understand the multinomial theorem from Section 1.7, one is led to symmetric polynomials and, in Section 1.8, to partitions of n. Elementary symmetric functions and their association with power sums lie at the

Combinatorics, *Second Edition*, by Russell Merris.
ISBN 0-471-26296-X

heart of Section 1.9. The final section of the chapter is an optional introduction to algorithms, the flavor of which can be sampled by venturing only as far as Algorithm 1.10.3. Those desiring not less but more attention to algorithms can find it in Appendix A2.

1.1. THE FUNDAMENTAL COUNTING PRINCIPLE



How many different four-letter words, including nonsense words, can be produced by rearranging the letters in LUCK? In the absence of a more inspired approach, there is always the brute-force strategy: Make a systematic list.

Once we become convinced that Fig. 1.1.1 accounts for every possible rearrangement and that no "word" is listed twice, the solution is obtained by counting the 24 words on the list.

While finding the brute-force strategy was effortless, implementing it required some work. Such an approach may be fine for an isolated problem, the *like* of which one does not expect to see again. But, just for the sake of argument, imagine yourself in the situation of having to solve a great many thinly disguised variations of this same problem. In that case, it would make sense to invest some effort in finding a strategy that requires less work to implement. Among the most powerful tools in this regard is the following commonsense principle.

1.1.1 Fundamental Counting Principle. Consider a (finite) sequence of decisions. Suppose the number of choices for each individual decision is independent of decisions made previously in the sequence. Then the number of ways to make the whole sequence of decisions is the product of these numbers of choices.

To state the principle symbolically, suppose c_i is the number of choices for decision i. If, for $1 \leq i < n$, c_{i+1} does not depend on which choices are made in

LUCK	LUKC	LCUK	LCKU	LKUC	LKCU
ULCK	ULKC	UCLK	UCKL	UKLC	UKCL
CLUK	CLKU	CULK	CUKL	CKLU	CKUL
KLUC	KLCU	KULC	KUCL	KCLU	KCUL

Figure 1.1.1. The rearrangements of LUCK.

decisions $1, \ldots, i$, then the number of different ways to make the sequence of decisions is $c_1 \times c_2 \times \cdots \times c_n$.

Let's apply this principle to the word problem we just solved. Imagine yourself in the midst of making the brute-force list. Writing down one of the words involves a sequence of four decisions. Decision 1 is which of the four letters to write first, so $c_1 = 4$. (It is no accident that Fig. 1.1.1 consists of four rows!) For each way of making decision 1, there are $c_2 = 3$ choices for decision 2, namely which letter to write second. Notice that the specific letters comprising these three choices depend on how decision 1 was made, but their *number* does not. That is what is meant by the number of choices for decision 2 being independent of how the previous decision is made. Of course, $c_3 = 2$, but what about c_4? Facing no alternative, is it correct to say there is "no choice" for the last decision? If that were literally true, then c_4 would be zero. In fact, $c_4 = 1$. So, by the fundamental counting principle, the number of ways to make the sequence of decisions, i.e., the number of words on the final list, is

$$c_1 \times c_2 \times c_3 \times c_4 = 4 \times 3 \times 2 \times 1.$$

The product $n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ is commonly written n! and read *n-factorial*.* The number of four-letter words that can be made up by rearranging the letters in the word LUCK is $4! = 24$.

What if the word had been LUCKY? The number of five-letter words that can be produced by rearranging the letters of the word LUCKY is $5! = 120$. A systematic list might consist of five rows each containing $4! = 24$ words.

Suppose the word had been LOOT? How many four-letter words, including nonsense words, can be constructed by rearranging the letters in LOOT? Why not apply the fundamental counting principle? Once again, imagine yourself in the midst of making a brute-force list. Writing down one of the words involves a sequence of four decisions. Decision 1 is which of the three letters L, O, or T to write first. This time, $c_1 = 3$. But, what about c_2? In this case, the number of choices for decision 2 depends on how decision 1 was made! If, e.g., *L* were chosen to be the first letter, then there would be two choices for the second letter, namely O or T. If, however, O were chosen first, then there would be three choices for the second decision, L, (the second) O, or T. Do we take $c_2 = 2$ or $c_2 = 3$? The answer is that *the fundamental counting principle does not apply to this problem* (at least not directly). The fundamental counting principle applies *only* when the *number* of choices for decision $i + 1$ is *independent* of how the previous i decisions are made.

To enumerate all possible rearrangements of the letters in LOOT, begin by distinguishing the two O's. maybe write the word as LOoT. Applying the fundamental counting principle, we find that there are $4! = 24$ different-*looking* four-letter words that can be made up from L, O, o, and T.

*The exclamation mark is used, not for emphasis, but because it is a convenient symbol common to most keyboards.

LOoT	LOTo	LoOT	LoTO	LTOo	LToO
OLoT	OLTo	OoLT	OoTL	OTLo	OToL
oLOT	oLTO	oOLT	oOTL	oTLO	oTOL
TLOo	TLoO	TOLo	TOoL	ToLO	ToOL

Figure 1.1.2. Rearrangements of LOoT.

Among the words in Fig. 1.1.2 are pairs like OLoT and oLOT, which look different only because the two O's have been distinguished. In fact, every word in the list occurs twice, once with "big O" coming before "little o", and once the other way around. Evidently, the number of different words (with indistinguishable O's) that can be produced from the letters in LOOT is not 4! but $4!/2 = 12$.

What about TOOT? First write it as TOot. Deduce that in any list of all possible rearrangements of the letters T, O, o, and t, there would be $4! = 24$ different-looking words. Dividing by 2 makes up for the fact that two of the letters are O's. Dividing by 2 again makes up for the two T's. The result, $24/(2 \times 2) = 6$, is the number of different words that can be made up by rearranging the letters in TOOT. Here they are

TTOO TOTO TOOT OTTO OTOT OOTT

All right, what if the word had been LULL? How many words can be produced by rearranging the letters in LULL? Is it too early to guess a pattern? Could the number we're looking for be $4!/3 = 8$? No. It is easy to see that the correct answer must be 4. Once the position of the letter U is known, the word is completely determined. Every other position is filled with an L. A complete list is ULLL, LULL, LLUL, LLLU.

To find out why 4!/3 is wrong, let's proceed as we did before. Begin by distinguishing the three L's, say L_1, L_2, and L_3. There are 4! different-looking words that can be made up by rearranging the four letters L_1, L_2, L_3, and U. If we were to make a list of these 24 words and then erase all the subscripts, how many times would, say, LLLU appear? The answer to this question can be obtained from the fundamental counting principle! There are three decisions: decision 1 has three choices, namely which of the three L's to write first. There are two choices for decision 2 (which of the two remaining L's to write second) and one choice for the third decision, which L to put last. Once the subscripts are erased, LLLU would appear 3! times on the list. We should divide $4! = 24$, not by 3, but by $3! = 6$. Indeed, $4!/3! = 4$ is the correct answer.

Whoops! if the answer corresponding to LULL is 4!/3!, why didn't we get 4!/2! for the answer to LOOT? In fact, we did: $2! = 2$.

Are you ready for MISSISSIPPI? It's the same problem! If the letters were all different, the answer would be 11!. Dividing 11! by 4! makes up for the fact that there are four I's. Dividing the quotient by another 4! compensates for the four S's.

Dividing that quotient by 2! makes up for the two P's. In fact, no harm is done if that quotient is divided by $1! = 1$ in honor of the single M. The result is

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

(Confirm the arithmetic.) The 11 letters in MISSISSIPPI can be (re)arranged in 34,650 different ways.*

There is a special notation that summarizes the solution to what we might call the "MISSISSIPPI problem."

1.1.2 Definition. The *multinomial coefficient*

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1!r_2!\cdots r_k!},$$

where $r_1 + r_2 + \cdots + r_k = n$.

So, "multinomial coefficient" is a *name* for the answer to the question, how many n-letter "words" can be assembled using r_1 copies of one letter, r_2 copies of a second (different) letter, r_3 copies of a third letter, ..., and r_k copies of a kth letter?

1.1.3 Example. After cancellation,

$$\begin{aligned}\binom{9}{4, 3, 1, 1} &= \frac{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 3 \times 2 \times 1 \times 1 \times 1} \\ &= 9 \times 8 \times 7 \times 5 = 2520.\end{aligned}$$

Therefore, 2520 different words can be manufactured by rearranging the nine letters in the word SASSAFRAS. □

In real-life applications, the words need not be assembled from the English alphabet. Consider, e.g., POSTNET† barcodes commonly attached to U.S. mail by the Postal Service. In this scheme, various numerical delivery codes‡ are represented by "words" whose letters, or *bits*, come from the alphabet $\left\{\mid, \middle|\right\}$. Corresponding, e.g., to a ZIP + 4 code is a 52-bit barcode that begins and ends with $\middle|$. The 50-bit middle part is partitioned into ten 5-bit zones. The first nine of these zones are for the digits that comprise the ZIP + 4 code. The last zone accommodates a *parity*

* This number is roughly equal to the number of members of the Mathematical Association of America (MAA), the largest professional organization for mathematicians in the United States.

† Postal Numeric Encoding Technique.

‡ The original five-digit Zoning Improvement Plan (ZIP) code was introduced in 1964; ZIP+4 codes followed about 25 years later. The 11-digit Delivery Point Barcode (DPBC) is a more recent variation.

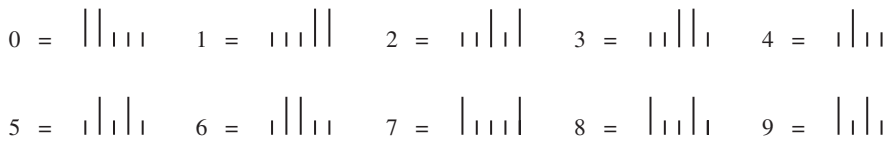


Figure 1.1.3. POSTNET barcodes.

check digit, chosen so that the sum of all ten digits is a multiple of 10. Finally, each digit is represented by one of the 5-bit barcodes in Fig. 1.1.3. Consider, e.g., the ZIP +4 code 20090-0973, for the Mathematical Association of America. Because the sum of these digits is 30, the parity check digit is 0. The corresponding 52-bit word can be found in Fig. 1.1.4.

20090-0973

Figure 1.1.4

We conclude this section with another application of the fundamental counting principle.

1.1.4 Example. Suppose you wanted to determine the number of positive integers that exactly divide $n = 12$. That isn't much of a problem; there are six of them, namely, 1, 2, 3, 4, 6, and 12. What about the analogous problem for $n = 360$ or for $n = 360,000$? Solving even the first of these by brute-force list making would be a lot of work. Having already found another strategy whose implementation requires a lot less work, let's take advantage of it.

Consider $360 = 2^3 \times 3^2 \times 5$, for example. If $360 = dq$ for positive integers d and q, then, by the uniqueness part of the *fundamental theorem of arithmetic*, the prime factors of d, together with the prime factors of q, are precisely the prime factors of 360, multiplicities included. It follows that the prime factorization of d must be of the form $d = 2^a \times 3^b \times 5^c$, where $0 \le a \le 3$, $0 \le b \le 2$, and $0 \le c \le 1$. Evidently, there are four choices for a (namely 0, 1, 2, or 3), three choices for b, and two choices for c. So, the number of possibile d's is $4 \times 3 \times 2 = 24$. □

1.1. EXERCISES

1 The Hawaiian alphabet consists of 12 letters, the vowels *a*, *e*, *i*, *o*, *u* and the consonants *h*, *k*, *l*, *m*, *n*, *p*, *w*.

(a) Show that 20,736 different 4-letter ''words'' could be constructed using the 12-letter Hawaiian alphabet.

(b) Show that 456,976 different 4-letter "words" could be produced using the 26-letter English alphabet.*

(c) How many four-letter "words" can be assembled using the Hawaiian alphabet if the second and last letters are vowels and the other 2 are consonants?

(d) How many four-letter "words" can be produced from the Hawaiian alphabet if the second and last letters are vowels but there are no restrictions on the other 2 letters?

2 Show that

(a) $3! \times 5! = 6!$.

(b) $6! \times 7! = 10!$.

(c) $(n+1) \times (n!) = (n+1)!$.

(d) $n^2 = n![1/(n-1)! + 1/(n-2)!]$.

(e) $n^3 = n![1/(n-1)! + 3/(n-2)! + 1/(n-3)!]$.

3 One brand of electric garage door opener permits the owner to select his or her own electronic "combination" by setting six different switches either in the "up" or the "down" position. How many different combinations are possible?

4 One generation back you have two ancestors, your (biological) parents. Two generations back you have four ancestors, your grandparents. Estimating 2^{10} as 10^3, approximately how many ancestors do you have

(a) 20 generations back?

(b) 40 generations back?

(c) In round numbers, what do you estimate is the total population of the planet?

(d) What's wrong?

5 Make a list of all the "words" that can be made up by rearranging the letters in

(a) TO. **(b)** TOO. **(c)** TWO.

6 Evaluate multinomial coefficient

(a) $\binom{6}{4,1,1}$. **(b)** $\binom{6}{3,3}$. **(c)** $\binom{6}{2,2,2}$.

*Based on these calculations, might it be reasonable to expect Hawaiian words, on average, to be longer than their English counterparts? Certainly such a conclusion would be warranted if both languages had the same vocabulary and both were equally "efficient" in avoiding long words when short ones are available. How efficient is English? Given that the total number of words defined in a typical "unabridged dictionary" is at most 350,000, one could, at least in principle, construct a new language with the same vocabulary as English but in which every word has four letters—and there would be 100,000 words to spare!

(d) $\binom{6}{3,2,1}$. **(e)** $\binom{6}{1,3,2}$. **(f)** $\binom{6}{1,1,1,1,1,1}$.

7 How many different "words" can be constructed by rearranging the letters in

(a) ALLELE? **(b)** BANANA? **(c)** PAPAYA?

(d) BUBBLE? **(e)** ALABAMA? **(f)** TENNESSEE?

(g) HALEAKALA? **(h)** KAMEHAMEHA? **(i)** MATHEMATICS?

8 Prove that

(a) $1 + 2 + 2^2 + 2^3 + \cdots + 2^n = 2^{n+1} - 1$.

(b) $1 \times 1! + 2 \times 2! + 3 \times 3! + \cdots + n \times n! = (n+1)! - 1$.

(c) $(2n)!/2^n$ is an integer.

9 Show that the barcodes in Fig. 1.1.3 comprise *all possible* five-letter words consisting of two |'s and three |'s.

10 Explain how the following barcodes fail the POSTNET standard:

(a) 

(b) 

(c) 

11 "Read" the ZIP+4 Code

(a) 

(b) 

12 Given that the first nine zones correspond to the ZIP+4 delivery code 94542-2520, determine the parity check digit and the two "hidden digits" in the 62-bit DPBC

(*Hint:* Do you need to be told that the parity check digit is last?)

13 Write out the 52-bit POSTNET barcode for 20742-2461, the ZIP+4 code at the University of Maryland used by the Association for Women in Mathematics.

14 Write out all 24 divisors of 360. (See Example 1.1.4.)

15 Compute the number of positive integer divisors of

(a) 2^{10}. **(b)** 10^{10}. **(c)** 12^{10}. **(d)** 31^{10}.

(e) 360,000. **(f)** 10!.

16 Prove that the positive integer n has an odd number of positive-integer divisors if and only if it is a perfect square.

17 Let $D = \{d_1, d_2, d_3, d_4\}$ and $R = \{r_1, r_2, r_3, r_4, r_5, r_6\}$. Compute the number

(a) of different functions $f : D \to R$.

(b) of one-to-one functions $f : D \to R$.

18 The latest automobile license plates issued by the California Department of Motor Vehicles begin with a single numeric digit, followed by three letters, followed by three more digits. How many different license "numbers" are available using this scheme?

19 One brand of padlocks uses combinations consisting of three (not necessarily different) numbers chosen from $\{0, 1, 2, \ldots, 39\}$. If it takes five seconds to "dial in" a three-number combination, how long would it take to try all possible combinations?

20 The *International Standard Book Number* (ISBN) is a 10-digit numerical code for identifying books. The groupings of the digits (by means of hyphens) varies from one book to another. The first grouping indicates where the book was published. In ISBN 0-88175-083-2, the zero shows that the book was published in the English-speaking world. The code for the Netherlands is "90" as, e.g., in ISBN 90-5699-078-0. Like POSTNET, ISBN employs a check digit scheme. The first nine digits (ignoring hyphens) are multiplied, respectively, by 10, 9, 8, ..., 2, and the resulting products summed to obtain S. In 0-88175-083-2, e.g.,

$$S = 10 \times 0 + 9 \times 8 + 8 \times 8 + 7 \times 1 + 6 \times 7 + 5 \times 5 + 4 \times 0 + 3 \times 8 + 2 \times 3 = 240.$$

The last (check) digit, L, is chosen so that $S + L$ is a multiple of 11. (In our example, $L = 2$ and $S + L = 242 = 11 \times 22$.)

(a) Show that, when S is divided by 11, the quotient Q and remainder R satisfy $S = 11Q + R$.

(b) Show that $L = 11 - R$. (When $R = 1$, the check digit is X.)

(c) What is the value of the check digit, L, in ISBN 0-534-95154-L?

(d) Unlike POSTNET, the more sophisticated ISBN system can not only detect common errors, it can sometimes "correct" them. Suppose, e.g., that a single digit is wrong in ISBN 90-5599-078-0. Assuming the check digit is correct, can you identify the position of the erroneous digit?

(e) Now that you know the position of the (single) erroneous digit in part (d), can you recover the correct ISBN?

(f) What if it were expected that exactly two digits were wrong in part (d). Which two digits might they be?

21 A total of $9! = 362,880$ different nine-letter "words" can be produced by rearranging the letters in FULBRIGHT. Of these, how many contain the four-letter sequence GRIT?

22 In how many different ways can eight coins be arranged on an 8×8 checkerboard so that no two coins lie in the same row or column?

23 If A is a finite set, its *cardinality*, $o(A)$, is the number of elements in A. Compute

(a) $o(A)$ when A is the set consisting of all five-digit integers, each digit of which is 1, 2, or 3.

(b) $o(B)$, where $B = \{x \in A : \text{each of } 1, 2, \text{and } 3 \text{ is among the digits of } x\}$ and A is the set in part (a).

1.2. PASCAL'S TRIANGLE

Mathematics is the art of giving the same name to different things.

— Henri Poincaré (1854–1912)

In how many different ways can an r-element subset be chosen from an n-element set S? Denote the number by $C(n, r)$. Pronounced "n-choose-r", $C(n, r)$ is just a name for the answer. Let's find the number represented by this name.

Some facts about $C(n, r)$ are clear right away, e.g., the nature of the elements of S is immaterial. All that matters is that there are n of them. Because the only way to choose an n-element subset from S is to choose all of its elements, $C(n, n) = 1$. Having n single elements, S has n single-element subsets, i.e., $C(n, 1) = n$. For essentially the same reason, $C(n, n-1) = n$: A subset of S that contains all but one element is uniquely determined by the one element that is left out. Indeed, this idea has a nice generalization. A subset of S that contains all but r elements is uniquely determined by the r elements that are left out. This natural one-to-one correspondence between subsets and their complements yields the following *symmetry property*:

$$C(n, n-r) = C(n, r).$$

1.2.1 Example. By definition, there are $C(5, 2)$ ways to select two elements from $\{A, B, C, D, E\}$. One of these corresponds to the two-element subset $\{A, B\}$. The complement of $\{A, B\}$ is $\{C, D, E\}$. This pair is listed first in the following one-to-one correspondence between two-element subsets and their three-element complements:

$$\{A,B\} \leftrightarrow \{C,D,E\}, \qquad \{B,D\} \leftrightarrow \{A,C,E\};$$
$$\{A,C\} \leftrightarrow \{B,D,E\}, \qquad \{B,E\} \leftrightarrow \{A,C,D\};$$
$$\{A,D\} \leftrightarrow \{B,C,E\}, \qquad \{C,D\} \leftrightarrow \{A,B,E\};$$
$$\{A,E\} \leftrightarrow \{B,C,D\}, \qquad \{C,E\} \leftrightarrow \{A,B,D\};$$
$$\{B,C\} \leftrightarrow \{A,D,E\}, \qquad \{D,E\} \leftrightarrow \{A,B,C\}.$$

By counting these pairs, we find that $C(5,2) = C(5,3) = 10$. □

A special case of symmetry is $C(n,0) = C(n,n) = 1$. Given n objects, there is just one way to reject all of them and, hence, just one way to choose none of them. What if $n = 0$? How many ways are there to choose no elements from the empty set? To avoid a deep philosophical discussion, let us simply adopt as a convention that $C(0,0) = 1$.

A less obvious fact about choosing these numbers is the following.

1.2.2 Theorem (Pascal's Relation). *If $1 \le r \le n$, then*

$$C(n+1,r) = C(n,r-1) + C(n,r). \tag{1.1}$$

Together with Example 1.2.1, Pascal's relation implies, e.g., that $C(6,3) = C(5,2) + C(5,3) = 20$.

Proof. Consider the $(n+1)$-element set $\{x_1, x_2, \ldots, x_n, y\}$. Its r-element subsets can be partitioned into two families, those that contain y and those that do not. To count the subsets that contain y, simply observe that the remaining $r-1$ elements can be chosen from $\{x_1, x_2, \ldots, x_n\}$ in $C(n,r-1)$ ways. The r-element subsets that do not contain y are precisely the r-element subsets of $\{x_1, x_2, \ldots, x_n\}$, of which there are $C(n,r)$. ■

The proof of Theorem 1.2.2 used another self-evident fact that is worth mentioning explicitly. (A much deeper extension of this result will be discussed in Chapter 2.)

1.2.3 The Second Counting Principle. If a set can be expressed as the disjoint union of two (or more) subsets, then the number of elements in the set is the sum of the numbers of elements in the subsets.

So far, we have been viewing $C(n,r)$ as a single number. There are some advantages to looking at these choosing numbers collectively, as in Fig. 1.2.1. The triangular shape of this array is a consequence of not bothering to write $0 = C(n,r)$, $r > n$. Filling in the entries we know, i.e., $C(n,0) = C(n,n) = 1$, $C(n,1) = n = C(n,n-1)$, $C(5,2) = C(5,3) = 10$, and $C(6,3) = 20$, we obtain Fig. 1.2.2.

n \ r	0	1	2	3	4	5	6	7
0	$C(0,0)$							
1	$C(1,0)$	$C(1,1)$						
2	$C(2,0)$	$C(2,1)$	$C(2,2)$					
3	$C(3,0)$	$C(3,1)$	$C(3,2)$	$C(3,3)$				
4	$C(4,0)$	$C(4,1)$	$C(4,2)$	$C(4,3)$	$C(4,4)$			
5	$C(5,0)$	$C(5,1)$	$C(5,2)$	$C(5,3)$	$C(5,4)$	$C(5,5)$		
6	$C(6,0)$	$C(6,1)$	$C(6,2)$	$C(6,3)$	$C(6,4)$	$C(6,5)$	$C(6,6)$	
7	$C(7,0)$	$C(7,1)$	$C(7,2)$	$C(7,3)$	$C(7,4)$	$C(7,5)$	$C(7,6)$	$C(7,7)$
				...				

Figure 1.2.1. $C(n, r)$.

Given the fourth row of the array (corresponding to $n = 3$), we can use Pascal's relation to compute $C(4, 2) = C(3, 1) + C(3, 2) = 3 + 3 = 6$. Similarly, $C(6, 4) = C(6, 2) = C(5, 1) + C(5, 2) = 5 + 10 = 15$. Continuing in this way, one row at a time, we can complete as much of the array as we like.

n \ r	0	1	2	3	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	$C(4,2)$	4	1			
5	1	5	10	10	5	1		
6	1	6	$C(6,2)$	20	$C(6,4)$	6	1	
7	1	7	$C(7,2)$	$C(7,3)$	$C(7,4)$	$C(7,5)$	7	1
				...				

Figure 1.2.2

Following Western tradition, we refer to the array in Fig. 1.2.3 as *Pascal's triangle.*[*] (Take care not to forget, e.g., that $C(6, 3) = 20$ appears, not in the third column of the sixth row, but in the fourth column of the seventh!)

Pascal's triangle is the source of many interesting identities. One of these concerns the sum of the entries in each row:

$$\begin{aligned} 1 + 1 &= 2, \\ 1 + 2 + 1 &= 4, \\ 1 + 3 + 3 + 1 &= 8, \\ 1 + 4 + 6 + 4 + 1 &= 16, \end{aligned} \tag{1.2}$$

[*]After Blaise Pascal (1623–1662), who described it in the book *Traité du triangle arithmétique*. Rumored to have been included in a lost mathematical work by Omar Khayyam (ca. 1050–1130), author of the *Rubaiyat*, the triangle is also found in surviving works by the Arab astronomer al-Tusi (1265), the Chinese mathematician Chu Shih-Chieh (1303), and the Hindu writer Narayana Pandita (1365). The first European author to mention it was Petrus Apianus (1495–1552), who put it on the title page of his 1527 book, *Rechnung*.

n \ r	0	1	2	3	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1
				⋯				

Figure 1.2.3. Pascal's triangle.

and so on. Why should each row sum to a power of 2? In

$$C(n,0) + C(n,1) + \cdots + C(n,n) = \sum_{r=0}^{n} C(n,r),$$

$C(n,0)$ is the number of subsets of $S = \{x_1, x_2, \ldots, x_n\}$ that have no elements; $C(n,1)$ is the number of one-element subsets of S; $C(n,2)$ is the number of two-element subsets, and so on. Evidently, the sum of the numbers in row n of Pascal's triangle is the total number of subsets of S (even when $n = 0$ and $S = \varnothing$). The empirical evidence from Equations (1.2) suggests that an n-element set has a total of 2^n subsets. How might one go about proving this conjecture?

One way to do it is by mathematical induction. There is, however, another approach that is both easier and more revealing. Imagine youself in the process of listing the subsets of $S = \{x_1, x_2, \ldots, x_n\}$. Specifying a subset involves a sequence of decisions. Decision 1 is whether to include x_1. There are two choices, *Yes* or *No*. Decision 2, whether to put x_2 into the subset, also has two choices. Indeed, there are two choices for each of the n decisions. So, by the fundamental counting principle, S has a total of $2 \times 2 \times \cdots \times 2 = 2^n$ subsets.

There is more. Suppose, for example, that $n = 9$. Consider the sequence of decisions that produces the subset $\{x_2, x_3, x_6, x_8\}$, a sequence that might be recorded as NYYNNYNYN. The first letter of this word corresponds to *No*, as in "no to x_1"; the second letter corresponds to *Yes*, as in "yes to x_2"; because x_3 is in the subset, the third letter is Y; and so on for each of the nine letters. Similarly, $\{x_1, x_2, x_3\}$ corresponds to the nine-leter word YYYNNNNNN. In general, there is a one-to-one correspondence between subsets of $\{x_1, x_2, \ldots, x_n\}$, and n-letter words assembled from the alphabet $\{\mathrm{N}, \mathrm{Y}\}$. Moreover, in this correspondence, r-element subsets correspond to words with r Y's and $n - r$ N's.

We seem to have discovered a new way to think about $C(n,r)$. It is the number of n-letter words that can be produced by (re)arranging r Y's and $n - r$ N's. This interpretation can be verified directly. An n-letter word consists of n spaces, or locations, occupied by letters. Each of the words we are discussing is completely determined once the r locations of the Y's have been chosen (the remaining $n - r$ spaces being occupied by N's).

The significance of this new perspective is that we know how to count the number of n-letter words with r Y's and $n-r$ N's. That's the MISSISSIPPI problem! The answer is multinomial coefficient $\binom{n}{r,n-r}$. Evidently,

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

For things to work out properly when $r = 0$ and $r = n$, we need to adopt another convention. Define $0! = 1$. (So, 0! is *not* equal to the nonsensical $0 \times (0-1) \times (0-2) \times \cdots \times 1$.)

It is common in the mathematical literature to write $\binom{n}{r}$ instead of $\binom{n}{r,n-r}$, one justification being that the information conveyed by "$n-r$" is redundant. It can be computed from n and r. The same thing could, of course, be said about *any* multinomial coefficient. The last number in the second row is always redundant. So, that particular argument is not especially compelling. The honest reason for writing $\binom{n}{r}$ is tradition.

We now have two ways to look at $C(n,r) = \binom{n}{r}$. One is what we might call the *combinatorial definition*: n-choose-r is the number of ways to choose r things from a collection of n things. The alternative, what we might call the *algebraic definition*, is

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

Don't make the mistake of asuming, just because it is more familiar, that the algebraic definition will always be easiest. (Try giving an algebraic proof of the identity $\sum_{r=0}^{n} C(n,r) = 2^n$.) Some applications are easier to approach using algebraic methods, while the combinatorial definition is easier for others. Only by becoming familiar with both will you be in a position to choose the easiest approach in every situation!

1.2.4 Example. In the basic version of poker, each player is dealt five cards (as in Fig. 1.2.4) from a standard 52-card deck (no joker). How many different five-card poker hands are there? Because someone (in a fair game it might be *Lady Luck*) chooses five cards from the deck, the answer is $C(52,5)$. The ways to find the number behind this name are: (1) Make an exhaustive list of all possible hands, (2) work out 52 rows of Pascal's triangle, or (3) use the algebraic definition

$$\begin{aligned} C(52,5) &= \frac{52!}{5!\,47!} \\ &= \frac{52 \times 51 \times 50 \times 49 \times 48 \times 47!}{5 \times 4 \times 3 \times 2 \times 1 \times 47!} \\ &= \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} \\ &= 52 \times 51 \times 10 \times 49 \times 2 \\ &= 2,598,960. \end{aligned}$$

□

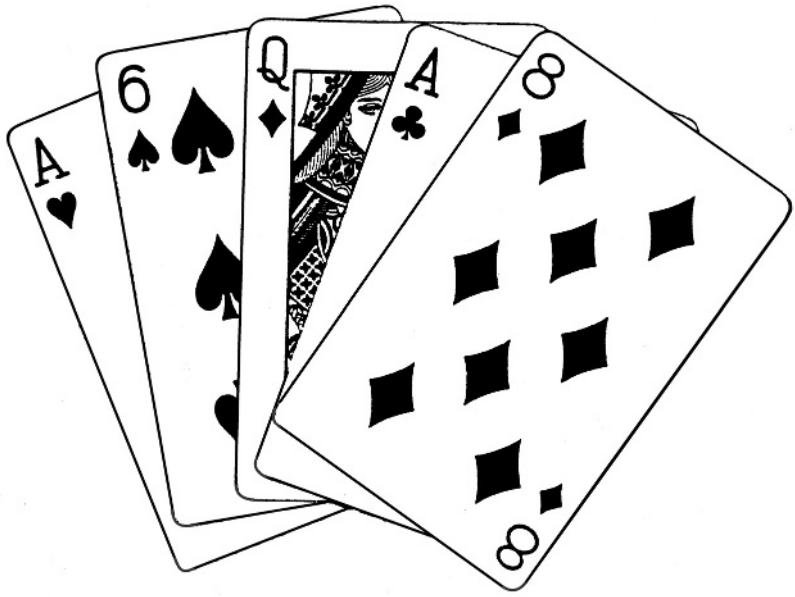


Figure 1.2.4. A five-card poker hand.

1.2.5 Example. The game of bridge uses the same 52 cards as poker.* The number of different 13-card bridge hands is

$$\begin{aligned} C(52,13) &= \frac{52!}{13!\,39!} \\ &= \frac{52 \times 51 \times \cdots \times 40 \times 39!}{13! \times 39!} \\ &= \frac{52 \times 51 \times \cdots \times 40}{13!}, \end{aligned}$$

about 635,000,000,000. □

It may surprise you to learn that $C(52,13)$ is so much larger than $C(52,5)$. On the other hand, it does seem clear from Fig. 1.2.3 that the numbers in each row of Pascal's triangle increase, from left to right, up to the middle of the row and then decrease from the middle to the right-hand end. Rows for which this property holds are said to be *unimodal*.

1.2.6 Theorem. *The rows of Pascal's triangle are unimodal.*

*The actual, physical cards are typically slimmer to accommodate the larger, 13-card hands.

Proof. If $n > 2r + 1$, the ratio

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

implying that $C(n, r+1) > C(n, r)$. ■

1.2. EXERCISES

1 Compute

(a) $C(7, 4)$. **(b)** $C(10, 5)$. **(c)** $C(12, 4)$.

(d) $C(101, 2)$. **(e)** $C(101, 99)$. **(f)** $C(12, 6)$.

2 If n and r are integers satisfying $n > r \geq 0$, prove that

(a) $(r+1)C(n, r+1) = (n-r)C(n, r)$.

(b) $(r+1)C(n, r+1) = nC(n-1, r)$.

3 Write out rows 7 through 10 of Pascal's triangle and confirm that the sum of the numbers in the 10th row is $2^{10} = 1024$.

4 Consider the sequence of numbers 0, 0, 1, 3, 6, 10, 15, ... from the third $(r = 2)$ column of Pascal's triangle. Starting with $n = 0$, the nth term of the sequence is $a_n = C(n, 2)$. Prove that, for all $n \geq 0$,

(a) $a_{n+1} - a_n = n$. **(b)** $a_{n+1} + a_n = n^2$.

5 Consider the sequence $b_0, b_1, b_2, b_3, \ldots$, where $b_n = C(n, 3)$. Prove that, for all $n \geq 0$,

(a) $b_{n+1} - b_n = C(n, 2)$.

(b) $b_{n+2} - b_n$ is a perfect square.

6 Poker is sometimes played with a joker. How many different five-card poker hands can be "chosen" from a deck of 53 cards?

7 Phrobana is a game played with a deck of 48 cards (no aces). How many different 12-card phrobana hands are there?

8 Give the inductive proof that an n-element set has 2^n subsets.

9 Let r_i be a positive integer, $1 \leq i \leq k$. If $n = r_1 + r_2 + \cdots + r_k$, prove that

$$\binom{n}{r_1, r_2, \ldots, r_k} = \binom{n-1}{r_1 - 1, r_2, \ldots, r_k} + \binom{n-1}{r_1, r_2 - 1, \ldots, r_k} + \cdots + \binom{n-1}{r_1, r_2, \ldots, r_k - 1}$$

(a) using algebraic arguments.
(b) using combinatorial arguments.

10 Suppose n, k, and r are integers that satisfy $n \geq k \geq r \geq 0$ and $k > 0$. Prove that

(a) $C(n,k)C(k,r) = C(n,r)C(n-r,k-r)$.

(b) $C(n,k)C(k,r) = C(n,k-r)C(n-k+r,r)$.

(c) $\sum_{j=0}^{n} C(n,j)C(j,r) = C(n,r)2^{n-r}$.

(d) $\sum_{j=k}^{n} (-1)^{j+k}C(n,j) = C(n-1,k-1)$.

11 Prove that $\left[\sum_{r=0}^{n} C(n,r)\right]^2 = \sum_{s=0}^{2n} C(2n,s)$.

12 Prove that $C(2n,n)$, $n > 0$, is always even.

13 Probably first studied by Leonhard Euler (1707–1783), the *Catalan sequence*[*] 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, ... is defined by $c_n = C(2n,n)/(n+1)$, $n \geq 0$. Confirm that the Catalan numbers satisfy

(a) $c_2 = 2c_1$. **(b)** $c_3 = 3c_2 - c_1$.

(c) $c_4 = 4c_3 - 3c_2$. **(d)** $c_5 = 5c_4 - 6c_3 + c_2$.

(e) $c_6 = 6c_5 - 10c_4 + 4c_3$. **(f)** $c_7 = 7c_6 - 15c_5 + 10c_4 - c_3$.

(g) Speculate about the general form of these equations.

(h) Prove or disprove your speculations from part (g).

14 Show that the Catalan numbers (Exercise 13) satisfy

(a) $c_n = C(2n-1,n-1) - C(2n-1,n+1)$.

(b) $c_n = C(2n,n) - C(2n,n-1)$.

(c) $c_{n+1} = \dfrac{4n+2}{n+2}c_n$.

15 One way to illustrate an r-element subset S of $\{1,2,\ldots,n\}$ is this: Let P_0 be the origin of the xy-plane. Setting $x_0 = y_0 = 0$, define

$$P_k = (x_k, y_k) = \begin{cases} (x_{k-1}+1, y_{k-1}) & \text{if} \quad k \in S, \\ (x_{k-1}, y_{k-1}+1) & \text{if} \quad k \notin S. \end{cases}$$

Finally, connect successive points by unit segments (either horizontal or vertical) to form a "path". Figure 1.2.5 illustrates the path corresponding to $S = \{3,4,6,8\}$ and $n = 8$.

[*]Euler was so prolific that more than one topic has come to be named for the first person to work on it *after* Euler, in this case, Eugene Catalan (1814–1894).

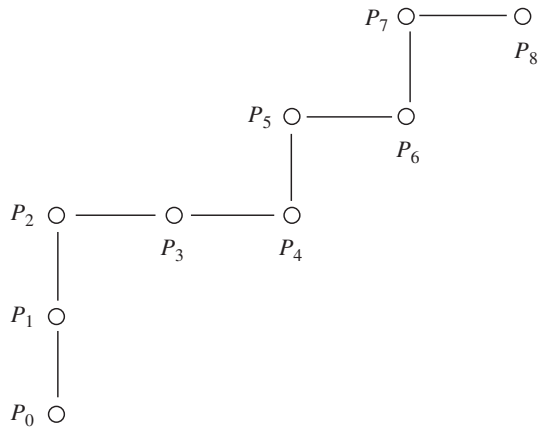


Figure 1.2.5

(a) Illustrate $E = \{2, 4, 6, 8\}$ when $n = 8$.

(b) Illustrate $E = \{2, 4, 6, 8\}$ when $n = 9$.

(c) Illustrate $D = \{1, 3, 5, 7\}$ when $n = 8$.

(d) Show that $P_n = (r, n - r)$ when S is an r-element set.

(e) A *lattice path* of *length n* in the xy-plane begins at the origin and consists of n unit "steps" each of which is either up or to the right. If r of the steps are to the right and $s = n - r$ of them are up, the lattice path terminates at the point (r, s). How many different lattice paths terminate at (r, s)?

16 Define $c_0 = 1$ and let c_n be the number of lattice paths of length $2n$ (Exercise 15) that terminate at (n, n) and never rise above the line $y = x$, i.e., such that $x_k \geq y_k$ for each point $P_k = (x_k, y_k)$. Show that

(a) $c_1 = 1$, $c_2 = 2$, and $c_3 = 5$.

(b) $c_{n+1} = \sum_{r=0}^{n} c_r c_{n-r}$. (*Hint:* Lattice paths "touch" the line $y = x$ for the last time at the point (n, n). Count those whose next-to-last touch is at the point (r, r)).

(c) c_n is the nth Catalan number of Exercises 13–14, $n \geq 1$.

17 Let X and Y be disjoint sets containing n and m elements, respectively. In how many different ways can an $(r + s)$-element subset Z be chosen from $X \cup Y$ if r of its elements must come from X and s of them from Y?

18 Packing for a vacation, a young man decides to take 3 long-sleeve shirts, 4 short-sleeve shirts, and 2 pairs of pants. If he owns 16 long-sleeve shirts, 20 short-sleeve shirts, and 13 pairs of pants, in how many different ways can he pack for the trip?

n \ r	0	1	2	3	4	5	6	7
0	$C(0,0)$							
1	$C(1,0)$	$C(1,1)$						
2	$C(2,0)$	$C(2,1)$	$C(2,2)$					
3	$C(3,0)$	$C(3,1)$	$C(3,2)$ +	**$C(3,3)$**				
4	$C(4,0)$	$C(4,1)$ +	**$C(4,2)$**	$C(4,3)$	$C(4,4)$			
5	$C(5,0)$ +	**$C(5,1)$**	$C(5,2)$	$C(5,3)$	$C(5,4)$	$C(5,5)$		
6	**$C(6,0)$**	$C(6,1)$	$C(6,2)$	$C(6,3)$	$C(6,4)$	$C(6,5)$	$C(6,6)$	
7	$C(7,0)$	$C(7,1)$	$C(7,2)$	$C(7,3)$	$C(7,4)$	$C(7,5)$	$C(7,6)$	$C(7,7)$
				⋯				

Figure 1.2.6

19 Suppose n is a positive integer and let $k = \lfloor n/2 \rfloor$, the greatest integer not larger than $n/2$. Define

$$F_n = C(n,0) + C(n-1,1) + C(n-2,2) + \cdots + C(n-k,k).$$

Starting with $n = 0$, the sequence $\{F_n\}$ is

$$1, 1, 2, 3, 5, 8, 13, \ldots,$$

where, e.g., the $7th$ number in the sequence, $F_6 = 13$, is computed by summing the **boldface** numbers in Fig. 1.2.6.*

(a) Compute F_7 directly from the definition.

(b) Prove the recurrence $F_{n+2} = F_{n+1} + F_n$, $n \geq 0$.

(c) Compute F_7 using part (b) and the initial fragment of the sequence given above.

(d) Prove that $\sum_{i=0}^{n} F_i = F_{n+2} - 1$.

20 C. A. Tovey used the Fibonacci sequence (Exercise 19) to prove that infinitely many pairs (n,k) solve the equation $C(n,k) = C(n-1,k+1)$. The first pair is $C(2,0) = C(1,1)$. Find the second. (*Hint*: $n < 20$. Your solution need not make use of the Fibonacci sequence.)

21 The Buda side of the Danube is hilly and suburban while the Pest side is flat and urban. In short, Budapest is a divided city. Following the creation of a new commission on culture, suppose 6 candidates from Pest and 4 from Buda volunteer to serve. In how many ways can the mayor choose a 5-member commission.

*It was the French number theorist François Édouard Anatole Lucas (1842–1891) who named these numbers after Leonardo of Pisa (ca. 1180–1250), a man also known as Fibonacci.

(a) from the 10 candidates?

(b) if proportional representation dictates that 3 members come from Pest and 2 from Buda?

22 H. B. Mann and D. Shanks discovered a criterion for primality in terms of Pascal's triangle: Shift each of the $n+1$ entries in row n to the right so that they begin in column $2n$. Circle the entries in row n that are multiples of n. Then r is prime if and only if all the entries in column r have been circled. Columns 0–11 are shown in Fig. 1.2.7. Continue the figure down to row 9 and out to column 20.

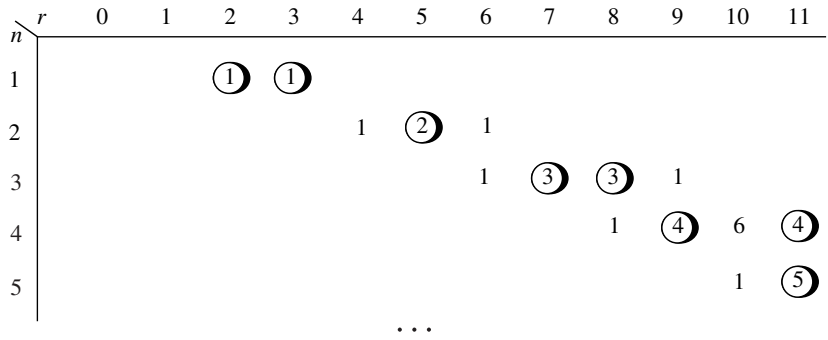


Figure 1.2.7

23 The superintendent of the Hardluck Elementary School District suggests that the Board of Education meet a \$5 million budget deficit by raising average class sizes, from 30 to 36 students, a 20% increase. A district teacher objects, pointing out that if the proposal is adopted, the potential for a *pair* of classmates to get into trouble will increase by 45%. What is the teacher talking about?

24 Strictly speaking, Theorem 1.2.6 establishes only half of the unimodality property. Prove the other half.

25 If n and r are nonnegative integers and x is an indeterminate, define $K(n,r) = (1+x)^n x^r$.

(a) Show that $K(n+1,r) = K(n,r) + K(n,r+1)$.

(b) Compare and contrast the identity in part (a) with Pascal's relation.

(c) Since part (a) is a polynomial identity, it holds when numbers are substituted for x. Let $k(n,r)$ be the value of $K(n,r)$ when $x=2$ and exhibit the numbers $k(n,r)$, $0 \le n$, $r \le 4$, in a 5×5 array, the rows of which are indexed by n and the columns by r. (*Hint:* Visually confirm that $k(n+1,r) = k(n,r) + k(n,r+1)$, $0 \le n$, $r \le 3$.)

26 Let S be an n-element set, where $n \geq 1$. If A is a subset of S, denote by $o(A)$ the *cardinality* of (number of elements in) A. Say that A is odd (even) if $o(A)$ is odd (even). Prove that the number of odd subsets of S is equal to the number of its even subsets.

27 Show that there are exactly seven different ways to factor $n = 63{,}000$ as a product of two relatively prime integers, each greater than one.

28 Suppose $n = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$, where $p_1, p_2, \ldots, p_r$ are distinct primes. Prove that there are exactly $2^{r-1} - 1$ different ways to factor n as a product of two relatively prime integers, each greater than one.

*1.3. ELEMENTARY PROBABILITY

> The theory of probabilities is basically only common sense reduced to calculation; it makes us appreciate with precision what reasonable minds feel by a kind of instinct, often being unable to account for it. ... It is remarkable that [this] science, which began with the consideration of games of chance, should have become the most important object of human knowledge.
>
> — Pierre Simon, Marquis de Laplace (1749–1827)

Elementary probability theory begins with the consideration of D equally likely "events" (or "outcomes"). If N of these are "noteworthy", then the probability of a noteworthy event is the fraction N/D. Maybe a brown paper bag contains a dozen jelly beans, say, 1 red, 2 orange, 2 blue, 3 green, and 4 purple. If a jelly bean is chosen at random from the bag, the probability that it will be blue is $\frac{2}{12} = \frac{1}{6}$; the probability that it will be green is $\frac{3}{12} = \frac{1}{4}$; the probability that it will be blue or green is $(2+3)/12 = \frac{5}{12}$; and the probability that it will be blue and green is $\frac{0}{12} = 0$.

Dice are commonly associated with games of chance. In a dice game, one is typically interested only in the numbers that rise to the top. If a single die is rolled, there are just six outcomes; if the die is "fair", each of them is equally likely. In computing the probability, say, of rolling a number greater than 4 with a single fair die, the denominator is $D = 6$. Since there are $N = 2$ noteworthy outcomes, namely 5 and 6, the probability we want is $P = \frac{2}{6} = \frac{1}{3}$.

The situation is more complicated when two dice are rolled. If all we care about is their sum, then there are 11 possible outcomes, anything from 2 to 12. But, the probability of rolling a sum, say, of 7 is not $\frac{1}{11}$ because these 11 outcomes are not equally likely. To help facilitate the discussion, assume that one of the dice is green and the other is red. Each time the dice are rolled, Lady Luck makes two decisions, choosing a number for the green die, and one for the red. Since there are 6 choices for each of them, the two decisions can be made in any one of $6^2 = 36$ ways. If both dice are fair, then *each of these* 36 *outcomes is equally likely.* Glancing at Fig. 1.3.1,

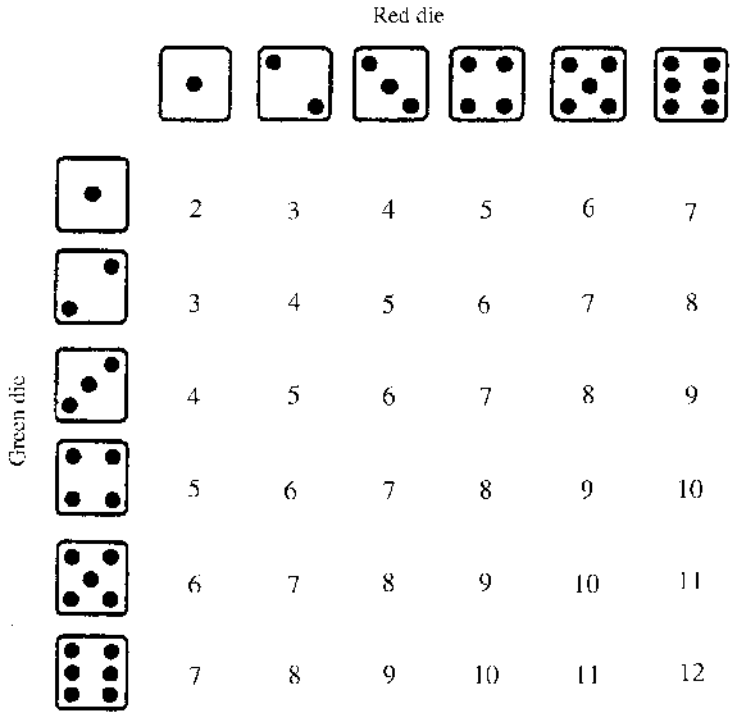


Figure 1.3.1. The 36 outcomes of rolling two dice.

one sees that there are six ways the dice can sum to 7, namely, a green 1 and a red 6, a green 2 and a red 5, a green 3 and a red 4, and so on. So, the probability of rolling a (sum of) 7 is not $\frac{1}{11}$ but $\frac{6}{36} = \frac{1}{6}$.

1.3.1 Example. Denote by $P(n)$ the probability of rolling (a sum of) n with two fair dice. Using Fig. 1.3.1, it is easy to see that $P(2) = \frac{1}{36} = P(12)$, $P(3) = \frac{2}{36} = \frac{1}{18} = P(11)$, $P(4) = \frac{3}{36} = \frac{1}{12} = P(10)$, and so on. What about $P(1)$? Since 1 is not among the outcomes, $P(1) = \frac{0}{36} = 0$. In fact, if P is some probability (any probability at all), then $0 \leq P \leq 1$. □

1.3.2 Example. A popular game at charity fundraisers is Chuck-a-Luck. The apparatus for the game consists of three dice housed in an hourglass-shaped cage. Once the patrons have placed their bets, the operator turns the cage and the dice roll to the bottom. If none of the dice comes up 1, the bets are lost. Otherwise, the operator matches, doubles, or triples each wager depending on the number of "aces" (1's) showing on the three dice.

Let's compute probabilities for various numbers of 1's. By the fundamental counting principle, there are $6^3 = 216$ possible outcomes (all of which are equally

Number of 1's	0	1	2	3
Probability	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Figure 1.3.2. Chuck-a-Luck probabilities.

likely if the dice are fair). Of these 216 outcomes, only one consists of three 1's. Thus, the probability that the bets will have to be tripled is $\frac{1}{216}$.

In how many ways can two 1's come up? Think of it as a sequence of two decisions. The first is which die should produce a number different from 1. The second is what number should appear on that die. There are three choices for the first decision and five for the second. So, there are $3 \times 5 = 15$ ways for the three dice to produce exactly two 1's. The probability that the bets will have to be doubled is $\frac{15}{216}$.

What about a single ace? This case can be approached as a sequence of three decisions. Decision 1 is which die should produce the 1 (three choices). The second decision is what number should appear on the second die (five choices, anything but 1). The third decision is the number on the third die (also five choices). Evidently, there are $3 \times 5 \times 5 = 75$ ways to get exactly one ace. So far, we have accounted for $1 + 15 + 75 = 91$ of the 216 possible outcomes. (In other words, the probability of getting *at least* one ace is $\frac{91}{216}$.) In the remaining $216 - 91 = 125$ outcomes, three are no 1's at all. These results are tabulated in Fig. 1.3.2. □

Some things, like determining which team kicks off to start a football game, are decided by tossing a coin. A fair coin is one in which each of the two possible outcomes, heads or tails, is equally likely. When a fair coin is tossed, the probability that it will come up heads is $\frac{1}{2}$.

Suppose four (fair) coins are tossed. What is the probability that half of them will be heads and half tails? Is it obvious that the answer is $\frac{3}{8}$? Once again, Lady Luck has a sequence of decisions to make, this time four of them. Since there are two choices for each decision, $D = 2^4$. With the noteworthies in **boldface**, these 16 outcomes are arrayed in Fig. 1.3.3. By inspection, $N = 6$, so the probability we seek is $\frac{6}{16} = \frac{3}{8}$.

HHHH	HTHH	THHH	**TTHH**
HHHT	**HTHT**	**THHT**	TTHT
HHTH	**HTTH**	**THTH**	TTTH
HHTT	HTTT	THTT	TTTT

Figure 1.3.3

1.3.3 Example. If 10 (fair) coins are tossed, what is the probability that half of them will be heads and half tails? Ten decisions, each with two choices, yields $D = 2^{10} = 1024$. To compute the numerator, imagine a systematic list analogous to Fig. 1.3.3. In the case of 10 coins, the noteworthy outcomes correspond to

10-letter "words" with five H's and five T's, so $N = \binom{10}{5,5} = C(10,5) = 252$, and the desired probability is $\frac{252}{1024} \doteq 0.246$. More generally, if n coins are tossed, the probability that exactly r of them will come up heads is $C(n,r)/2^n$.

What about the probability that *at most* r of them will come up heads? That's easy enough: $P = N/2^n$, where $N = N(n,r) = C(n,0) + C(n,1) + \cdots + C(n,r)$ is the number of n-letter words that can be assembled from the alphabet $\{H,T\}$ and that contain at most r H's. □

Here is a different kind of problem: Suppose two fair coins are tossed, say a dime and a quarter. If you are told (only) that one of them is heads, what is the probability that the other one is also heads? (Don't just guess, think about it.)

May we assume, without loss of generality, that the dime is heads? If so, because the quarter has a head of its own, so to speak, the answer should be $\frac{1}{2}$. To see why this is wrong, consider the equally likely outcomes when two fair coins are tossed, namely, HH, HT, TH, and TT. If all we know is that one (at least) of the coins is heads, then TT is eliminated. Since the remaining three possibilities are still equally likely, $D = 3$, and the answer is $\frac{1}{3}$.

There are two "morals" here. One is that the most reliable guide to navigating probability theory is *equal likelihood*. The other is that finding a correct answer often depends on having a precise understanding of the question, and that requires precise language.

1.3.4 Definition. A nonempty finite set E of equally likely outcomes is called a *sample space*. The number of elements in E is denoted $o(E)$. For any subset A of E, the probability of A is $P(A) = o(A)/o(E)$. If B is a subset of E, then $P(A \text{ or } B) = P(A \cup B)$, and $P(A \text{ and } B) = P(A \cap B)$.

In mathematical writing, an unqualified "or" is inclusive, as in "A or B or both".*

1.3.5 Theorem. *Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then*

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Proof. The sum $o(A) + o(B)$ counts all the elements of A and all the elements of B. It even counts some elements twice, namely those in $A \cap B$. Subtracting $o(A \cap B)$ compensates for this double counting and yields

$$o(A \cup B) = o(A) + o(B) - o(A \cap B).$$

(Notice that this formula generalizes the second counting principle; it is a special case of the even more general principle of inclusion and exclusion, to be discussed in Chapter 2.) It remains to divide both sides by $o(E)$ and use Definition 1.3.4.

■

*The exclusive "or" can be expressed using phrases like "either A or B" or "A or B but not both".

1.3.6 Corollary. *Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then $P(A \text{ or } B) \leq P(A) + P(B)$ with equality if and only if A and B are disjoint.*

Proof. $P(A \text{ and } B) = 0$ if and only if $o(A \cap B) = 0$ if and only if $A \cap B = \varnothing$. ■

A special case of this corollary involves the *complement*, $A^c = \{x \in E : x \notin A\}$. Since $A \cup A^c = E$ and $A \cap A^c = \varnothing$, $o(A) + o(A^c) = o(E)$. Dividing both sides of this equation by $o(E)$ yields the useful identity

$$P(A) + P(A^c) = 1.$$

1.3.7 Example. Suppose two fair dice are rolled, say a red one and a green one. What is the probability of rolling a 3 on the red die, call it a red 3, or a green 2? Let's abbreviate by setting $R3 = \text{red } 3$ and $G2 = \text{green } 2$ so that, e.g., $P(R3) = \frac{1}{6} = P(G2)$.

Solution 1: When both dice are rolled, only one of the $6^2 = 36$ equally likely outcomes corresponds to $R3$ and $G2$, so $P(R3 \text{ and } G2) = \frac{1}{36}$. Thus, by Theorem 1.3.5,

$$\begin{aligned} P(R3 \text{ or } G2) &= P(R3) + P(G2) - P(R3 \text{ and } G2) \\ &= \tfrac{1}{6} + \tfrac{1}{6} - \tfrac{1}{36} \\ &= \tfrac{11}{36}. \end{aligned}$$

Solution 2: Let P_c be the complementary probability that neither $R3$ nor $G2$ occurs. Then $P_c = N/D$, where $D = 36$. The evaluation of N can be viewed in terms of a sequence of two decisions. There are five choices for the "red" decision, anything but number 3, and five for the "green" one, anything but number 2. Hence, $N = 5 \times 5 = 25$, and $P_c = \frac{25}{36}$, so the probability we want is

$$P(R3 \text{ or } G2) = 1 - P_c = \tfrac{11}{36}.$$

□

1.3.8 Example. Suppose a single (fair) die is rolled twice. What is the probability that the first roll is a 3 or the second roll is a 2? Solution: $\frac{11}{36}$. This problem is equivalent to the one in Example 1.3.7. □

1.3.9 Example. Suppose a single (fair) die is rolled twice. What is the probability of getting a 3 or a 2?

Solution 1: Of the $6 \times 6 = 36$ equally likely outcomes, $4 \times 4 = 16$ involve neither a 3 nor a 2. The complementary probability is $P(2 \text{ or } 3) = 1 - \frac{16}{36} = \frac{5}{9}$.

Solution 2: There are two ways to roll a 3 *and* a 2; either the 3 comes first followed by the 2 or the other way around. So, $P(3 \text{ and } 2) = \frac{2}{36} = \frac{1}{18}$. Using Theorem 1.3.5, $P(3 \text{ or } 2) = \frac{1}{6} + \frac{1}{6} - \frac{1}{18} = \frac{5}{18}$.

Whoops! Since $\frac{5}{9} \neq \frac{5}{18}$, one (at least) of these "solutions" is incorrect. The probability computed in solution 1 is greater than $\frac{1}{2}$, which *seems* too large. On the other hand, it is not hard to spot an error in solution 2, namely, the incorrect application of Theorem 1.3.5. The calculation $P(3) = \frac{1}{6}$ would be valid had the die been rolled only *once*. For this problem, the correct interpretation of $P(3)$ is the probability that the first roll is 3 or the second roll is 3. That should be identical to the probability determined in Example 1.3.8. (Why?) Using the (correct) values $P(3) = P(2) = \frac{11}{36}$ in solution 2, we obtain $P(2 \text{ or } 3) = \frac{11}{36} + \frac{11}{36} - \frac{1}{18} = \frac{5}{9}$.

The next time you get a chance, roll a couple of dice and see if you can avoid both 2's and 3's more than 44 times out of 99. □

Another approach to $P(A \text{ and } B)$ emerges from the notion of "conditional probability".

1.3.10 Definition. Let E be a fixed but arbitrary sample space. If A and B are subsets of E, the *conditional probability*

$$P(B|A) = \begin{cases} P(B) & \text{if } A = \varnothing, \\ o(A \cap B)/o(A) & \text{otherwise.} \end{cases}$$

When A is not empty, $P(B|A)$ may be viewed as the probability of B given that A is certain (e.g., known already to have occurred). The problem of tossing two fair coins, a dime and a quarter, involved conditional probabilities. If h and t represent heads and tails, respectively, for the dime and H and T for the quarter, then the sample space $E = \{hH, hT, tH, tT\}$. If $A = \{hH, hT, tH\}$ and $B = \{hH\}$, then $P(B|A) = \frac{1}{3}$ is the probability that both coins are heads given that one of them is. If $C = \{hH, hT\}$, then $P(B|C) = \frac{1}{2}$ is the probability that both coins are heads given that the dime is.

1.3.11 Theorem. *Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then*

$$P(A \text{ and } B) = P(A)P(B|A).$$

Proof. Let $D = o(E)$, $a = o(A)$, and $N = o(A \cap B)$. If $a = 0$, there is nothing to prove. Otherwise, $P(A) = a/D$, $P(B|A) = N/a$, and $P(A)P(B|A) = (a/D)(N/a) = N/D = P(A \text{ and } B)$. ■

1.3.12 Corollary (Bayes's* First Rule). *Let E be a fixed but arbitrary sample space. If A and B are subsets of E, then $P(A)P(B|A) = P(B)P(A|B)$.*

Proof. Because $P(A \text{ and } B) = P(B \text{ and } A)$, the result is immediate from Theorem 1.3.11. ■

1.3.13 Definition. Suppose E is a fixed but arbitrary sample space. Let A and B be subsets of E. If $P(B|A) = P(B)$, then A and B are *independent*.

Definitions like this one are meant to associate a name with a phenomenon. In particular, Definition 1.3.13 is to be understood in the sense that A and B are independent if *and only if* $P(B|A) = P(B)$. (In statements of theorems, on the other hand, "if" should never be interpreted to mean "if and only if".)

In plain English, A and B are independent if $A = \varnothing$ or if $A \neq \varnothing$ and the probability of B is the same whether A is known to have occurred or not. It follows from Corollary 1.3.12 (and the definition) that $P(B|A) = P(B)$ if and only if $P(A|B) = P(A)$, i.e., A and B are independent if and only if B and A are independent. A combination of Definition 1.3.13 and Theorem 1.3.11 yields

$$P(A \text{ and } B) = P(A)P(B) \tag{1.3}$$

if and only if A and B are independent.

Equation (1.3) is analogous to the case of equality in Corollary 1.3.6, i.e., that

$$P(A \text{ or } B) = P(A) + P(B) \tag{1.4}$$

if and only if A and B are disjoint. Let's compare and contrast the words *independent* and *disjoint*.

1.3.14 Example. Suppose a card is drawn from a standard 52-card deck. Let K represent the outcome that the card is a king and C the outcome that it is a club.[†] Because $P(C) = \frac{13}{52} = \frac{1}{4} = P(C|K)$, these outcomes are independent and, as expected,

$$\begin{aligned} P(K)P(C) &= \left(\tfrac{1}{13}\right)\left(\tfrac{1}{4}\right) \\ &= \tfrac{1}{52} \\ &= P(\text{king of clubs}) \\ &= P(K \text{ and } C). \end{aligned}$$

Because $K \cap C = \{\text{king of clubs}\} \neq \varnothing$, K and C are not disjoint. As expected, $P(K \text{ or } C) = \frac{16}{52}$ differs from $P(K) + P(C) = \frac{4}{52} + \frac{13}{52} = \frac{17}{52}$ by $\frac{1}{52} = P(K \text{ and } C)$.

If Q is the outcome that the card is a queen, then K and Q are disjoint but not independent. In particular, $P(K \text{ or } Q) = \frac{8}{52} = P(K) + P(Q)$, but $P(Q) = \frac{4}{52} = \frac{1}{13}$ while $P(Q|K) = 0$.

[*]Thomas Bayes (1702–1761), an English mathematician and clergyman, was among those who defended Newton's calculus against the philosophical attack of Bishop Berkeley. He is better known, however, for his *Essay Towards Solving a Problem in the Doctrine of Chances*.

[†]Alternatively, let E be the set of all 52 cards, K the four-element subset of kings, and C the subset of all 13 clubs.

Finally, let F be the outcome that the chosen card is a "face card" (a king, queen, or jack). Because $K \cap F = K \neq \varnothing$, outcomes K and F are not disjoint. Since $P(F) = \frac{12}{52} = \frac{3}{13}$ while $P(F|K) = 1$, neither are they independent. □

1.3.15 Example. Imagine two copy editors independently proofreading the same manuscript. Suppose editor X finds x typographical errors while editor Y finds y. Denote by z the number of typos discovered by both editors so that, together, they identify a total of $x + y - z$ errors. George Pólya showed* how this information can be used to estimate the number of typographical errors overlooked by both editors! If the manuscript contains a total of t typos, then the empirical probability that editor X discovered (some randomly chosen) one of them is $P(X) = x/t$. If, on the other hand, one of the errors discovered by Y is chosen at random, the empirical probability that X found it is $P(X|Y) = z/y$. If X is a consistent, experienced worker, these two "productivity ratings" should be about the same. Setting $z/y \doteq x/t$ (i.e., assuming $P(X|Y) \doteq P(X)$) yields $t \doteq xy/z$. □

1.3.16 Example. In the popular game *Yahtzee*, five dice are rolled in hopes of obtaining various outcomes. Suppose you needed to roll three 4's to win the game. What is the probability of rolling exactly three 4's in a single throw of the five dice?

Solution: There are $C(5, 3) = 10$ ways for Lady Luck to choose three dice to be the 4's, e.g., the "first" three dice might be 4's while the remaining two are not; dice 1, 2, and 5 might be 4's while dice 3 and 4 are not; and so on. Label these ten outcomes $A_1, A_2, \ldots, A_{10}$.

The computation of $P(A_1)$, say, is a classic application of Equation (1.3). The probability of rolling a 4 on one die is independent of the number rolled on any of the other dice. Since the probability that any one of the first three dice shows a 4 is $\frac{1}{6}$ and the probability that either one of the last two does not is $\frac{5}{6}$,

$$P(A_1) = \tfrac{1}{6} \times \tfrac{1}{6} \times \tfrac{1}{6} \times \tfrac{5}{6} \times \tfrac{5}{6}.$$

Similarly, $P(A_i) = \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$, $2 \leq i \leq 10$.

If, e.g.,

$$A_1 = \{\text{dice } 1, 2, \text{ and } 3 \text{ are 4's while dice 4 and 5 are not}\}$$

and

$$A_3 = \{\text{dice } 1, 2, \text{ and } 5 \text{ are 4's while dice 3 and 4 are not}\},$$

*In a 1976 article published in the *American Mathematical Monthly.*

then the third die is a 4 in every outcome belonging to A_1 while it is anything but a 4 in each outcome of A_3, i.e., $A_1 \cap A_3 = \varnothing$. Similarly, A_i and A_j are disjoint for all $i \neq j$. Therefore, by Equation (1.4),

$$\begin{aligned} P(\text{three 4's}) &= P(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_{10}) \\ &= P(A_1) + P(A_2) + \cdots + P(A_{10}) \\ &= 10\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^2. \end{aligned}$$

So, the probability of rolling exactly three 4's in a single throw of five fair dice is

$$C(5,3)\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^2 = 0.032\cdots . \qquad \square$$

Example 1.3.16 illustrates a more general pattern. The probability of rolling exactly r 4's in a single throw of n fair dice is $C(n,r)\left(\frac{1}{6}\right)^r\left(\frac{5}{6}\right)^{n-r}$. If a single fair die is thrown n times, the probability of rolling exactly r 4's is the same: $C(n,r)\left(\frac{1}{6}\right)^r\left(\frac{5}{6}\right)^{n-r}$. A similar argument applies to n independent attempts to perform any other "trick". If the probability of a successful attempt is p, then the probability of an unsuccessful attempt is $q = 1 - p$, and the probability of being successful in exactly r of the n attempts is

$$P(r) = C(n,r)p^r q^{n-r}, \qquad 0 \leq r \leq n. \tag{1.5}$$

Equation (1.5) governs what has come to be known as a *binomial probability distribution*.

1.3. EXERCISES

1 According to an old adage, it is unsafe to eat shellfish during a month whose name does not contain the letter *R*. What is the probability that it is unsafe to eat shellfish (according to the adage) during a randomly chosen month of the year?

2 Suppose two fair dice are rolled. What is the probability that their sum is

(a) 5? **(b)** 6? **(c)** 8? **(d)** 9?

3 Suppose three fair dice are rolled. What is the probability that their sum is

(a) 5? **(b)** 9? **(c)** 12? **(d)** 15?

4 Suppose a fair coin is tossed 10 times and the result is 10 successive heads. What is the probability that heads will be the outcome the next time the coin is tossed? (If you didn't know the coin was fair, you might begin to suspect otherwise. The *chi-squared* statistic, which is beyond the scope of this book, affords a way to estimate the probability that a fair coin would produce discrepancies from expected behavior that are this bad or worse.)

5 Many game stores carry *dodecahedral* dice having 12 pentagonal faces numbered 1–12. Suppose a pair of fair dodecahedral dice are rolled. What is the probability that they will sum to

(a) 5? **(b)** 7? **(c)** 13? **(d)** 25?

6 In what fraction of six-child families are half the children girls and half boys? (Assume that boys and girls are equally likely.)

7 Suppose you learn that in a particular two-child family one (at least) of the children is a boy. What is the probability that the other child is a boy? (Assume that boys and girls are equally likely.)

8 Suppose the king and queen of hearts are shuffled together with the king and queen of spades and all four cards are placed face down on a table.

(a) If your roommate picks up two of the cards and says, "I have a king!" what is the probability that s/he has both kings? (Don't just guess. Work it out as if your life depended on getting the right answer.)

(b) If your roommate picks up two of the cards and says, "I have the king of spades", what is the probability that s/he has both kings?

9 In the Chuck-a-Luck game of Example 1.3.2, show how the fundamental counting principle can be used to enumerate the outcomes that don't contain any 1's at all.

10 Suppose that six dice are tossed. What is the probability of rolling exactly

(a) three 4's? **(b)** four 4's? **(c)** five 4's?

11 Suppose that five cards are chosen at random from a standard 52-card deck. Show that the probability they comprise a "flush" is about $\frac{1}{505}$. (A flush is a poker hand each card of which comes from the same suit.)

12 Suppose some game of chance offers the possibility of winning one of a variety of prizes. Maybe there are n prizes with values $v_1, v_2, \ldots, v_n$. If the probability of winning the ith prizes is p_i, then the *expected value* of the game is

$$\sum_{i=1}^{n} v_i P_i.$$

Consider, e.g., a version of Chuck-a-Luck in which, on any given turn, you win \$1 for each ace.

(a) Show that the expected value of this game is 50 cents. (*Hint:* Figure 1.3.2.)

(b) What is the maximum amount anyone should be willing to pay for the privilege of playing this version each time the cage is turned?

(c) What is the maximum amount anyone should be willing to wager on this version each time the cage is turned? (The difference between "paying for the privilege of playing" and "wagering" is that, in the first case, your payment is lost, regardless of the outcome, whereas in the second case, you keep your wager unless the outcome is no aces at all.)

13 Does Chuck-a-Luck follow a binomial probability distribution? (Justify your answer.)

14 Suppose four fair coins are tossed. Let A be the set of outcomes in which at least two of the coins are heads, B the set in which at most two of the coins are heads, and C the set in which exactly two of the coins are heads. Compute

(a) $P(A)$. **(b)** $P(B)$. **(c)** $P(C)$.
(d) $P(A|B)$. **(d)** $P(A|C)$. **(e)** $P(A \text{ or } B)$.

15 In 1654, Antoine Gombaud, the Chevalier de Méré, played a game in which he bet that at least one 6 would result when four dice are rolled. What is the probability that de Méré won in any particular instance of this game? (Assume the dice were fair.)

16 Perhaps beause he could no longer find anyone to take his bets (see Exercise 15), the Chevalier de Méré switched to betting that, in any 24 consecutive rolls of two (fair) dice, "boxcars" (double 6's) would occur at least once. What is the probability that he won in any particular instance of this new game?

17 Suppose you toss a half-dollar coin n times. How large must n be to guarantee that your probability of getting heads at least once is better than 0.99?

18 The following problem was once posed by the diarist Samuel Pepys to Isaac Newton. "Who has the greatest chance of success: a man who throws six dice in hopes of obtaining at least one 6; a man who throws twelve dice in hopes of obtaining at least two 6's; or a man who throws eighteen dice in hopes of obtaining at least three 6's?" Compute the probability of success in each of the three cases posed by Pepys.

19 Are $P(A|B)$ and $P(B|A)$ always the same? (Justify your answer.)

20 Suppose that each of k people secretly chooses as integer between 1 and m (inclusive). Let P be the probability that some two of them choose the same number. Compute P (rounded to two decimal places) when

(a) $(m,k) = (10,4)$ **(b)** $(m,k) = (20,6)$ **(c)** $(m,k) = (365,23)$

(*Hint:* Compute the complementary probability that everyone chooses different numbers.)

21 Suppose 23 people are chosen at random from a crowd. Show the probability that some two of them share the same birthday (just the day, not the day and year) is greater than $\frac{1}{2}$. (Assume that none of them was born on February 29.)

22 Let E be a fixed but arbitrary sample space. Let A and B be nonempty subsets of E. Prove that A and B cannot be both independent and disjoint.

23 The four alternate die numberings illustrated in Fig. 1.3.4 were discovered by Stanford statistician Bradley Efron. Note that when dice A and B are thrown together, die A beats (rolls a higher number than) die B with probability $\frac{2}{3}$. Compute the probability that

(a) die B beats die C.

(b) die C beats die D.

(c) die D beats die A.

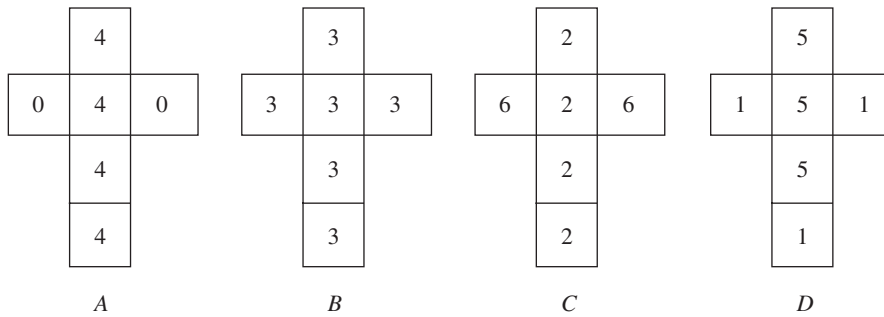


Figure 1.3.4. Efron dice.

24 One variation on the notion of a *random walk* takes place in the first quadrant of the xy-plane. Starting from the origin, the direction of each "step" is determined by the flip of a coin. If the kth coin flip is "heads", the kth step is one unit in the positive x-direction; if the coin flip is "tails", the step is one unit in the positive y-direction.

(a) Show that, after n steps, a random walker arrives at a point $P_n = (r, n - r)$, where $n \geq r \geq 0$. (*Hint:* Exercise 15, Section 1.2.)

(b) Assuming the coin is fair, compute the probability that the point $P_8 = (4, 4)$.

(c) Assuming the coin is fair, compute the probability that P_7 lies on the line $y = x$.

(d) Assuming the coin is fair, compute the probability that P_{2k} lies on the line $y = x$.

(e) Let r and n be fixed integers, $n \geq r \geq 0$. Assuming the coin comes up heads with probability p and tails with probability $q = 1 - p$, compute the probability that, after n steps, a random walker arrives at the point $P_n = (r, n - r)$.

25 Imagine having been bitten by an exotic, poisonous snake. Suppose the ER physician estimates that the probability you will die is $\frac{1}{3}$ unless you receive effective treatment immediately. At the moment, she can offer you a choice of experimental antivenins from two competing "snake farms." Antivenin X has been administered to ten previous victims of the same type of snake bite and nine of them survived. Antivenin Y, on the other hand, has only been administered to four previous patients, but all of them survived. Unfortunately, mixing the two drugs in your body would create a toxic substance much deadlier than the venom from the snake. Under these circumstances, which antivenin would you choose, and why?

26 In California's *SuperLotto Plus* drawing of February 16, 2002, three winners shared a record $193 million jackpot. *SuperLotto Plus* players choose five numbers, ranging from 1 through 47 *Plus* a "Mega" number between 1 and 27 (inclusive). The winning numbers in the drawing of February 16 were 6, 11, 31, 32, and 39 *Plus* 20. (Order matters only to the extent that the Mega number is separate from the other five numbers.)

(a) Compute the probability of winning a share of the jackpot (with a single ticket).

(b) The jackpot is not the only prize awarded in the *SuperLotto* game. In the February 16 drawing, 56 tickets won $27,859 (each) by matching all five (ordinary) numbers but missing the Mega number. Compute the probability of correctly guessing all five (ordinary) numbers.

(c) Compute the probability of correctly guessing all five (ordinary) numbers and missing the Mega number.

(d) In the February 16 drawing, 496 ticket holders won $1572 (each) by correctly guessing the Mega number and four out of the other five. Compute the probability of winning this prize (with a single ticket).

*1.4. ERROR-CORRECTING CODES

```
0  0  0  0  0  0  0  0  0  0  0  0
  1  1  1  1  1  1  1  1  1  1  1
0  0  0  0  0  1  1  1  1  1  1  0
  1  1  1  1  1  0  0  0  0  0  0
0  1  0  1  0  1  0  1  0  1  0  0
  1  0  1  0  1  0  1  0  1  0  1
0  0  0  0  0  0  0  0  0  0  0  0
  1  1  1  1  1  1  1  1  1  1  1
0  1  0  1  0  1  0  1  0  1  0  0
  1  0  1  0  1  0  1  0  1  0  1
1  1  1  1  1  1  1  1  1  1  1  1
```

The key to the connection between the combinatorial and algebraic definitions of $C(n, r) = \binom{n}{r}$ involves n-letter words constructed from two-letter alphabets. A binary code is a vocabulary comprised of such words. Binary codes have a wide

variety of applications ranging from stunning interplanetary images to everyday digital recordings. A common theme in these applications is the reliable movement of data through unreliable communication channels. The general problem is to detect and correct transmission errors that might arise from something as mundane as scratches on a CD to something as exotic as solar flares during an interplanetary voyage.

Our primary focus will be on words assembled using the alphabet $F = \{0, 1\}$, the letters of which are typically called *bits*.

1.4.1 Definition. An n-bit word is also known as a *binary word* of *length n*. The set of all 2^n binary words of length n will be denoted F^n. A *binary code* of *length n* is a nonempty subset of F^n.

A "good" code is one that can be used to transmit lots of information down a noisy channel, quickly and reliably. Consider, e.g., the code $\mathscr{C} = \{00000, 11111\} \subset F^5$, where 00000 might represent "yes" and 11111 might mean "no." Suppose one of these two codewords is sent down a noisy channel, only to have 000_0, or worse, 00010 come out the other end. While it is a binary word of length 5, 00010 is not a codeword. Thus, we *detect* an error. Just to make things interesting, suppose no further communication is possible. (Maybe the original message consisted of a single prerecorded burst.) Assuming it is more likely for any particular bit to be transmitted correctly than not, 00000 is more likely to have been the transmitted message than 11111. Thus, we might *correct* 00010 to 00000. Note that a binary word "corrected" in this way need not be correct in the sense that it was the transmitted codeword. It is just the legitimate codeword most likely to be correct.

1.4.2 Definition. Suppose b and w are binary words of length n. The *distance* between them, $d(b, w)$, is the number of places in which they differ.

Nearest-neighbor decoding refers to a process by which an erroneous binary word w is corrected to a legitimate codeword c in a way that minimizes $d(w, c)$. With the code $\mathscr{C} = \{00000, 11111\}$, it is possible to detect as many as four errors. With nearest-neighbor decoding, it is possible (correctly) to correct as many as two; $\mathscr{C}$ is a two-*error-correcting code*. (If 00000 were sent and 10101 received, nearest-neighbor decoding would produce 11111, the wrong message, Code $\mathscr{C}$ is not three-error correcting.)

1.4.3 Definition. An *r-error-correcting code* is one for which nearest-neighbor decoding reliably corrects as many as r errors.

Using the code $\mathscr{C} = \{100, 101\}$, suppose 100 is sent. If 110 is received, an error is detected. Because $d(110, 100) = 1 < 2 = d(110, 101)$ nearest-neighbor decoding corrects 110 to 100, the correct message. But, this is not enough to make $\mathscr{C}$ a one-error-correcting code. If 100 is sent and a single transmission error occurs, in the third bit, so that 101 is received, the error will not even be detected, much

less corrected. An r-error-correcting code must reliably correct r erroneous bits, no matter which r bits they happen to be.

Calling d a "distance" doesn't make it one. To be a distance , $d(b, w)$ should be zero whenever $b = w$, positive whenever $b \neq w$, symmetric in the sense that $d(b, w) = d(w, b)$ for all b and w, and it should satisfy the shortest-distance-between-two-points rule, also known as the *triangle inequality*. Of these conditions, only the last one is not obviously valid.

1.4.4 Lemma (Triangle Inequality). *If u, v, and w are fixed but arbitrary binary words of length n, then*

$$d(u, w) \leq d(u, v) + d(v, w).$$

Proof. The words u and w cannot differ from each other in a place where neither of them differs from v. Being binary words, they also cannot differ from each other in a place where both of them differ from v. It follows that $d(u, w)$ is the sum of the number of places where u differs from v but w does not, and the number of places where w differs from v but u does not. Because the first term in this sum is at most $d(u, v)$, the number of places where u differs from v, and the second is at most $d(w, v)$, the number of places where w differs from v, $d(u, w) \leq d(u, v) + d(w, v)$. ■

1.4.5 Definition. An (n, M, d) *code* consists of M binary words of length n, the minimum distance between any pair of which is d.

1.4.6 Example. The code $\{00000, 11111\}$, is evidently a (5, 2, 5) code. While it is easy to see that $n = 5$ and $M = 4$ for the code $\mathscr{C} = \{00000, 11101, 10011, 01110\}$, the value of d is less obvious. Computing the distances $d(00000, 11101) = 4$, $d(00000, 10011) = 3$, $d(00000, 01110) = 3$, $d(11101, 10011) = 3$, $d(11101, 01110) = 3$, and $d(10011, 01110) = 4$, between all $C(4, 2) = 6$ pairs of codewords, yields the minimum $d = 3$. So, $\mathscr{C}$ is a (5, 4, 3) code. □

An (n, M, d) code $\mathscr{C}$ can reliably detect as many as $d - 1$ errors. To determine how many erros $\mathscr{C}$ can reliably correct, consider the possibility that, for some erroneous binary word w, there is a tie for the codeword nearest w. Maybe $d(c, w) \geq r$ for every $c \in \mathscr{C}$, with equality for c_1 and c_2. In practice, such ties are broken by some predetermined rule. Because it can happen that this arbitrary rule dictates decoding w as c_1, even when c_2 was the transmitted codeword, no such code can *reliably* correct as many as r errors. However, by the triangle inequality, $d(c_1, w) = d(w, c_2) = r$ implies that $d(c_1, c_2) \leq 2r$, guaranteeing that no such situation can occur when $2r < d$. It seems we have proved the following.

1.4.7 Theorem. *An (n, M, d) code is r-error-correcting if and only if $2r + 1 \leq d$.*

Recall that our informal notion of a good code is one that can transmit lots of information down a noisy channel, quickly and reliably. So far, our discussion has focused on reliability. Let's talk about speed. For the sake of rapid transmission, one would like to have short words (small n) and a large vocabulary (big M). Because $M \leq 2^n$, these are conflicting requirements.

Suppose we fix n and d and ask how large M can be. The following notion is useful in addressing this question.

1.4.8 Definition. Let w be a binary word of length n. The *sphere* of *radius r centered at w* is

$$S_r(w) = \{b \in F^n : d(w, b) \leq r\},$$

the set of binary words that differ from w in at most r bits.

Because it is a sphere together with its interior, "ball" might be a more appropriate name for $S_r(w)$.

1.4.9 Example. Let $\mathscr{C}$ be a $(10, M, 7)$ code and suppose $c \in \mathscr{C}$. Because there are 10 places in which a binary word can differ from c, there must be 10 binary words that differ from c in just 1 place. Similarly, $C(10, 2) = 45$ words differ from c in exactly 2 places and $C(10, 3) = 120$ words differ from it in 3 places. Evidently, including c itself, $S_3(c)$ contains a total of

$$1 + 10 + 45 + 120 = 176$$

binary words only one of which, namely, c, is a codeword.

If c_1 and c_2 are different codewords, then $S_3(c_1) \cap S_3(c_2) \neq \varnothing$ only if there is a binary word w such that $d(w, c_1) \leq 3$ and $d(w, c_2) \leq 3$, implying that

$$\begin{aligned} d(c_1, c_2) &\leq d(c_1, w) + d(w, c_2) \\ &\leq 6 \end{aligned}$$

and contradicting our assumption that the minimum distance between codewords is 7. In other words, if $c_1 \neq c_2$, then $S_3(c_1) \cap S_3(c_2) = \varnothing$.

One might think of $S_3(c)$ as a *sphere of influence* for c. Because different spheres of influence are disjoint and since each sphere contains 176 of the 1024 binary words of length 10, there is insufficient room in F^{10} for as many as six spheres of influence. (Check it: $6 \times 176 = 1056$.) Evidently, the vocabulary of a three-error-correcting binary code of length 10 can consist of no more than five words! If $\mathscr{C}$ is a (10, M, 7) code, then $M \leq 5$. □

Example 1.4.9 has the following natural generalization.

1.4.10 Theorem (Sphere-Packing Bound). *The vocabulary of an r-error-correcting code of length n contains no more than* $2^n / N(n, r)$ *codewords, where*

$$N(n, r) = C(n, 0) + C(n, 1) + \cdots + C(n, r).$$

Proof. Suppose $\mathscr{C} = \{c_1, c_2, \ldots, c_M\} \subset F^n$ is an r-error-correcting code. Let $S_r(c_i)$ be the sphere of influence centered at codeword c_i, $1 \leq i \leq M$. Since spheres corresponding to different codewords are disjoint and $o(S_r(c_i)) = N(n, r)$, $1 \leq i \leq M$, the number of different binary words of length n contained in the union of the M spheres is $M \times N(n, r)$, a number that cannot exceed the total number of binary words of length n. ■

1.4.11 Example. Suppose you were asked to design a three-error-correcting code capable of sending the four messages NORTH, EAST, WEST, or SOUTH. Among the easiest solutions is the (16, 4, 8) code

$$\{0000000000000000, 1111111100000000, 1111000011110000, 1111000000001111\}.$$

However, if speed (or professional pride) is an issue, you might want to hold this one in reserve and look for something better.

For a solution to be optmal, it should (at the very least) be an $(n, 4, 7)$ code with n as small as possible. According to Example 1.4.9, a three-error-correcting code of length 10 can have at most five codewords, which would be ample for our needs. Moreover, because $4 \times N(9, 3) = 4 \times (1 + 9 + 36 + 84) = 520 > 2^9$, there can be no (9, 4, 7) codes. So, the best we can hope to achieve is a (10, 4, 7) code.

Without loss of generality, we can choose $c_1 = 0000000000$. (Why?) Since it must differ from c_1 in (no fewer than) 7 places, we may as well let $c_2 = 1111111000$. To differ from c_1 in 7 places, c_3 must contain 7 (or more) 1's. But, c_3 can differ from c_2 in 7 places only if (at least) four of its first seven bits are 0's! It is, of course, asking too much of a 10-bit word that it contain at least four 0's and at least seven 1's. The same problem arises no matter which seven bits are set equal to 1 in c_2, and setting more than seven bits equal to 1 only makes matters worse! It seems there do not exist even three binary words of length 10 each differing from the other two in (at least) seven bits. (Evidently, the sphere-packing bound is not always attainable!)

If there are no (10, 3, 7) codes, there certainly cannot be any (10, 4, 7) codes. What about an (11, 4, 7) code? This time, the obvious choices, $c_1 = 00000000000$ and $c_2 = 11111110000$, leave room for $c_3 = 00001111111$, which differs from c_2 in eight places and from c_1 in seven. Because $c_4 = 11110001111$ differs from c_2 and c_3 in seven places and from c_1 in eight, $\mathscr{C} = \{c_1, c_2, c_3, c_4\}$ is an (11, 4, 7) code. □

Our discovery, in Example 1.4.11, that $M \leq 2$ in any $(10, M, 7)$ code is a little surprising. Because a sphere of radius 3 in F^{10} holds (only) 176 words, two nonoverlapping spheres contain little more than a third of the 1024 words in F^{10}! On the other hand, how many solid Euclidean balls of radius 3 will fit inside a Euclidean cube of volume 1024?*

*Even in the familiar world of three-dimensional Euclidean space, sphere-packing problems can be highly nontrivial. On the other hand, in at least one sense, packing Euclidean spheres in three-space is a bad analogy. Orange growers are interested in sphere packing because, without damaging the produce, they want to minimize the fraction of empty space in each "full" box of oranges. Apart from degenerate cases, equality is *never* achievable in the grower's version of the sphere-packing bound.

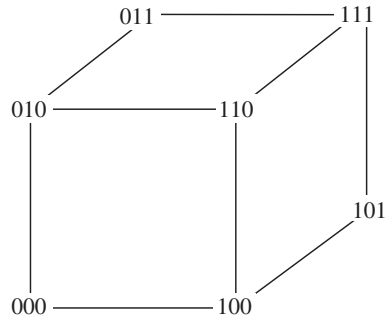


Figure 1.4.1. Three-dimensional binary space.

1.4.12 Example. As illustrated in Fig. 1.4.1, three-dimensional binary space F^3 is comparable, not to a Euclidean cube, but to the set consisting of its eight vertices! While packing the Euclidean cube with Euclidean spheres always results in "left-over" Euclidean points, F^3 is easily seen* to be the disjoint union of the spheres $S_1(000) = \{000, 100, 010, 001\}$ and $S_1(111) = \{111, 011, 101, 110\}$. (Note the two different ways in which $S_1(111)$ is "complementary" to $S_1(000)$.) □

1.4.13 Definition. An (n, M, d) code is *perfect* if $2^n = M \times [C(n,0) + C(n,1) + \cdots + C(n,r)]$, where $r = \lfloor (d-1)/2 \rfloor$ is the greatest integer not exceeding $(d-1)/2$.

So, an r-error-correcting code $\mathscr{C}$ is perfect if and only if its vocabulary achieves the sphere-packing bound, if and only if F^n is the disjoint union of the spheres $S_r(c)$ as c ranges over $\mathscr{C}$, if and only if every binary word of length n belongs to the sphere of influence of some (unique) codeword. In particular, a perfect code is as *efficient* as it is possible for codes to be.

It follows from Definition 1.4.13 that F^n, itself, is perfect. It is the disjoint union of the (degenerate) spheres $S_0(b)$, $b \in F^n$. Such trivial examples are uninteresting for a number of reasons, not the least of which is that F^n cannot detect, much less correct, even a single error. A nontrivial perfect code emerges from Example 1.4.12, namely, the one-error-correcting $(3, 2, 3)$ code $\{000, 111\}$. Might this be the only nontrivial example? No, $\{100, 011\}$ is another. All right, might the only nontrivial examples have parameters $(3, 2, 3)$?

1.4.14 Lemma. *Suppose $\mathscr{C}$ is an (n, M, d) code for which $r = \lfloor (d-1)/2 \rfloor = 1$. Then $\mathscr{C}$ is perfect if and only if there exists an integer $m \geq 2$ such that $n = 2^m - 1$ and $M = 2^{n-m}$.*

Proof. If $\mathscr{C}$ is perfect, then $2^n = M \times N(n,1) = M(1+n)$, so that $M = 2^n/(1+n)$. Now, $1+n$ exactly divides 2^n only if $1+n = 2^m$ for some positive integer

*Because one vertex is hidden from view, "seen" may not be the most appropriate word to use here.

$m \leq n$, in which case $M = 2^n/2^m = 2^{n-m}$. Moreover, $2^m - 1 = n \geq d \geq 3$ implies $m \geq 2$.

Conversely, if $n = 2^m - 1$ and $M = 2^{n-m}$, then $M(1+n) = 2^{n-m} \times 2^m = 2^n$. ■

1.4.15 Example. The parameters of the perfect (3, 2, 3) code $\mathscr{C} = \{000, 111\}$ satisfy the conditions of Lemma 1.4.14 when $m = 2$.

Setting $d = 3$ and $m = 3$ in Lemma 1.4.14 shows that every (7, 16, 3) code is perfect. What it does not show is the existence of even one (7, 16, 3) code! However, as the reader may confirm, (7, 16, 3) is the triple of parameters for the so-called *Hamming code* $\mathscr{H}_3 = \{0000000, 1000011, 0100101, 0010110, 0001111, 1100110, 1010101, 1001100, 0110011, 0101010, 0011001, 0111100, 1011010, 1101001, 1110000, 1111111\}$. In Chapter 6, the existence of an $(n, M, 3)$ code that satisfies the conditions of Lemma 1.4.14 will be established for every $m \geq 4$. □

1.4. EXERCISES

1 What is the largest possible value for M in any $(8, M, 1)$ code?

2 How many errors can an $(n, M, 8)$ code

(a) detect? **(b)** correct?

3 Find the parameters (n, M, d) for the binary code

(a) $\mathscr{C}_1 = \{000, 011, 101, 110\}$.

(b) $\mathscr{C}_2 = \{000, 011, 101, 110, 111, 100, 010, 001\}$.

(c) $\mathscr{C}_3 = \{0000, 0110, 1010, 1100, 1111, 1001, 0101, 0011\}$.

(d) $\mathscr{C}_4 = \{11000, 00011, 00101, 00110, 01001, 01010, 01100, 10001, 10010, 10100\}$, (Compare $\mathscr{C}_4$ with the POSTNET barcodes of Fig. 1.1.3.)

4 Construct a code (or explain why none exists) with parameters

(a) (3, 4, 2). **(b)** (6, 4, 4). **(c)** (12, 4, 8).

(d) (4, 7, 2). **(e)** (8, 7, 4). **(f)** (8, 8, 4).

5 The *American Standard Code for Information Interchange* (ASCII) is a scheme for assigning numerical values from 0 through 255 to selected symbols. For example, the uppercase letters of the English alphabet correspond to 65 through 90, respectively. Why 256 symbols? Good question. The answer involves bits and bytes. Consisting of two four-bit "zones", a *byte* can store any binary numeral in the range 0 through 255.

Apart from representing binary numerals, bytes can also be viewed as codewords in $\mathscr{C} = F^8$. Because it corresponds to the base-2 numeral for 65, the codeword/byte 01000001 represents *A* (in the ASCII scheme). Similarly, *Z*, corresponding to 90, is represented by the codeword/byte 01011010.

(a) What is the ASCII number for the letter *S*?

(b) What byte represents *S*?

(c) What letter corresponds to ASCII number 76?

(d) What letter is represented by codeword/byte 01010101?

(e) The ASCII number for the square-root symbol is 251. What codeword/byte represents $\surd$?

(f) Decode the message 01001101-01000001-01010100-01001000.

6 The *complement* of a binary word b is the word b^* obtained from b by changing all if its zeros to ones and all of its ones to zeros. For any binary code $\mathscr{C}$, define $\mathscr{C}^* = \{c^* : c \in \mathscr{C}\}$.

(a) Show that $\mathscr{C}_2 = \mathscr{C}_1 \cup \mathscr{C}_1^*$, where $\mathscr{C}_1$ and $\mathscr{C}_2$ are the codes in Exercises 3(a) and (b), respectively.

(b) Find a code $\mathscr{C}$ of length 3 satisfying $\mathscr{C}^* = F^3 \backslash \mathscr{C}$, the set-theoretic complement of $\mathscr{C}$. (*Hint:* Example 1.4.12.)

(c) Find a code $\mathscr{C}$ of length 3 satisfying $\mathscr{C}^* = \mathscr{C}$.

(d) If $\mathscr{C}$ is an (n, M, d) code, prove or disprove that $\mathscr{C}^*$ has the same parameters.

(e) If $\mathscr{C}$ is an (n, M, d) code, prove or disprove that $F^n \backslash \mathscr{C}$ has the same parameters.

7 The *weight* of a binary word b, $\mathrm{wt}(b)$, is the number of bits of b equal to 1. A *constant-weight code* is one in which every codeword has the same weight.

(a) Show that $d \geq 2$ in any constant-weight (n, M, d) code (in which $n > 1$).

(b) Find a constant-weight $(8, M, d)$ code with $d > 2$.

(c) Find the largest possible value for M in a constant-weight (8, M, d) code.

8 Let $\mathscr{C}$ be the (8, 56, 2) code consisting of all binary words of length 8 and weight 5. (See Exercise 7.) Let $\mathscr{C}^*$ be the code consisting of the complements of the codewords of $\mathscr{C}$. (See Exercise 6.) Prove that $\mathscr{C} \cup \mathscr{C}^*$ is an (8, 112, 2) code.

9 M. Plotkin* proved that if $n < 2d$ in the (n, M, d) code $\mathscr{C}$, then $M \leq 2\lfloor d/(2d - n) \rfloor$, where $\lfloor x \rfloor$ is the greatest integer not larger that x. Does the Plotkin bound preclude the existence of

(a) (12, 52, 5) codes? **(b)** (12, 7, 7) codes?

(c) (13, 13, 7) codes? **(d)** (15, 2048, 3) codes?

(Justify your answers.)

10 Does the sphere-packing bound (Theorem 1.4.10) rule out the existence of a

(a) (12, 52, 5) code? **(b)** (12, 7, 7) code?

(c) (13, 13, 7) code? **(d)** (15, 2048, 3) code?

(Justify your answers.)

*Binary codes with specified minimum distances, *IEEE Trans. Info. Theory* 6 (1960), 445–450.

11 The purpose of this exercise is to prove the Plotkin bound from Exercise 9. Let $\mathscr{C} = \{c_1, c_2, \ldots, c_M\}$ be an (n, M, d) code where $n < 2d$. Define

$$D = \sum_{i,j=1}^{M} d(c_i, c_j).$$

(a) Prove that $D \geq M(M-1)d$.

(b) Let A be the $M \times n$ (0, 1)-matrix whose ith row consists of the bits of codeword c_i. If the kth column of A contains z_k 0's (and $M - z_k$ 1's), prove that

$$D = 2\sum_{k=1}^{n} z_k(M - z_k).$$

(c) If M is even, show that $f(z) = z(M-z)$ is maximized when $z = \frac{1}{2}M$.
(d) Prove the Plotkin bound in the case that M is even.
(e) If M is odd, show that $D \leq \frac{1}{2}n(M^2 - 1)$.
(f) Prove the Plotkin bound in the case that M is odd.
(g) Where is the hypothesis $n < 2d$ used in the proof?

12 The *parity* of binary word b is 0 if $\text{wt}(b)$ is even and 1 if $\text{wt}(b)$ is odd. (See Exercise 7.) If $b = xy\ldots z$ is a binary word of length n and parity p, denote by $b^+ = xy\ldots zp$ the binary word of length $n+1$ obtained from b by appending a new bit equal to its parity. For any binary code $\mathscr{C}$ of length n, let $\mathscr{C}^+ = \{c^+ : c \in \mathscr{C}\}$.

(a) Show that $\mathscr{C}_3 = \mathscr{C}_2^+$, where $\mathscr{C}_2$ and $\mathscr{C}_3$ are the codes from Exercises 3(b) and (c), respectively.
(b) If $\mathscr{C}$ is an (n, M, d) code, where d is odd, prove that $\mathscr{C}^+$ is an $(n+1, M, d+1)$ code.
(c) Prove that exactly half the words in F^n have parity $p = 0$.
(d) Prove or disprove that if $\mathscr{C}$ is a fixed but arbitrary binary code of length n, then exactly half the words in $\mathscr{C}$ have even weight.

13 Let $M(n, d)$ be the largest possible value of M in any (n, M, d) code. Prove that $M(n, 2r-1) = M(n+1, 2r)$.

14 If $\mathscr{C}$ is a code of length n, its "weight enumerator" is the two-variable polynomial defined by

$$W_{\mathscr{C}}(x, y) = \sum_{c \in \mathscr{C}} x^{\text{wt}(c)} y^{n - \text{wt}(c)},$$

where $\text{wt}(c)$ is the weight of c defined in Exercise 7.

(a) Compute $W_{\mathscr{C}}(x, y)$ for each of the codes in Exercise 3.

(b) Show that $W_{\mathscr{C}}(x, y) = x^7 + 7x^4y^3 + 7x^3y^4 + y^7$ for the perfect Hamming code $\mathscr{C} = \mathscr{H}_3$ of Example 1.4.15.

(c) Two codes are *equivalent* if one can be obtained from the other by uniformly permuting (rearranging) the order of the bits in each codeword. Show that equivalent codes have the same parameters.

(d) Show that equivalent codes have the same weight enumerator.

(e) Exhibit two inequivalent codes with the same weight enumerator.

15 Exhibit the parameters for the perfect Hamming code $\mathscr{H}_4$ (corresponding to $m = 4$ in Lemma 1.4.14).

16 Show that the Plotkin bound (Exercise 9) is strong enough to preclude the existence of a (10, 3, 7) code (see Example 1.4.11).

17 Can the (11, 4, 7) code in Example 1.4.11 be extended to an (11, 5, 7) code?

18 Let u, v, and w be binary words of length n. Show that $d(u, w) = d(u, v) + d(v, w) - 2b$, where b is the number of places in which u and w both differ from v.

19 Following up on the discussion between Examples 1.4.11 and 1.4.12, show that two solid Euclidean spheres of radius 3 cannot be fit inside a cubical box of volume 1024 in such a way that both spheres touch the bottom of the box.

20 Show that the necessary condition for the existence of an r-error-correcting code given by the sphere-packing bound is not sufficient.

21 Let $M(n, d)$ be the largest possible value of M in any (n, M, d) code.

(a) If $n \geq 2$, prove that $M(n, d) \leq 2M(n - 1, d)$.

(b) Prove that $M(2d, d) \leq 4d$.

22 Show that a necessary condition for equality to hold in the Plotkin bound (Exercises 9 and 11) is $d(c_i, c_j) = d$, $i \neq j$.

23 The (7, 16, 3) code $\mathscr{H}_3$ in Example 1.4.15 is advertised as a perfect code. While it is easy to check that $\mathscr{H}_3$ is a binary code of length 7 containing 16 codewords, (given what we know now) it might take a minite or two to confirm that the minimum distance between any two codewords is 3. Assuming that has been done, how hard is it to confirm that $\mathscr{H}_3$ is a perfect code? (Justify your answer by providing the confirmation.)

24 Let $A = F^3 \backslash S_1(110)$ the (set-theoretic) complement of $S_1(110)$ in F^3.

(a) Show that A is a sphere in F^3.

(b) Do A and $S_1(110)$ exhibit both kinds of complementarity discussed in Exercise 6?

25 Prove that every (23, 4096, 7) code is perfect.

26 Construct a code with parameters (8, 16, 4).

27 Construct a code with parameters

(a) (6, 8, 3). **(b)** (7, 8, 4).

28 The purpose of this exercise is to justify nearest-neighbor decoding. We begin with some assumptions about the transmission channel. The simplest case is a so-called *symmetric* channel in which the probability of a 1 being changed to 0 is the same as that of a 0 being changed to 1. If we assume this common error probability, call it p, is the same for each bit of every word, then $q = 1 - p$ is the probability that any particular bit is transmitted correctly.

(a) Show that the probability of transmitting codeword c and receiving binary word w along such a channel is $p^r q^{n-r}$, where r is the number of places in which c and w differ.

(b) Under the assumption that $p < \frac{1}{2}$ (engineers work very hard to ensure that p is *much* less than $\frac{1}{2}$, show that the probability in part (a) is maximized when r is as small as possible.

29 Suppose the two-error-correcting code $\mathscr{C} = \{00000, 11111\}$ is used in a symmetric channel for which the probability of a transmission error in each bit is $p = 0.05$. (See exercise 28.)

(a) Show that the probability of more than two errors in the transmission of a single codeword is less than 0.0012.

(b) There may be cases in which a probability of failure as high as 0.0012 is unacceptable. What is the probability of more than three errors in the transmission of a single codeword using the same channel and the code $\{0000000, 1111111\}$?

1.5. COMBINATORIAL IDENTITIES

Poetry is the art of giving different names to the same thing.

— Anonymous

As we saw in Section1.2, $C(n, r) = \binom{n}{r}$ is the same as multinomial coefficient $\binom{n}{r, n-r}$. In fact, $C(n, r)$ is commonly called a *binomial* coefficient.* Given that binomial coefficients are special cases of multinomial coefficients, it is natural to wonder whether we still need a separate name and notation for n-choose-r. On the other hand, it turns out that multinomial coefficients can be expressed as products of binomial coefficients. Thus, one could just as well argue for discarding the multinomial coefficients!

1.5.1 Theorem. *If $r_1 + r_2 + \cdots + r_k = n$, then*

$$\binom{n}{r_1, r_2, \ldots, r_k} = \binom{n}{r_1}\binom{n - r_1}{r_2}\binom{n - r_1 - r_2}{r_3} \cdots \binom{n - r_1 - r_2 - \cdots - r_{k-1}}{r_k}.$$

*This name is thought to have been coined by Michael Stifel (ca. 1485–1567), among the most celebrated algebraists of the sixteenth century. Also known for numerological prophesy, Stifel predicted publicly that the world would end on October 3, 1533.

Proof. Multinomial coefficient $\binom{n}{r_1, r_2, \ldots, r_k}$ is the number of n-letter "words" that can be assembled using r_1 copies of one "letter", say A_1; r_2 copies of a second, A_2; and so on, finally using r_k copies of some kth character, A_k. The theorem is proved by counting these words another way and setting the two (different-looking) answers equal to each other.

Think of the process of writing one of the words as a sequence of k decisions. Decision 1 is which of n spaces to fill with A_1's. Because this amounts to selecting r_1 of the n available positions, it involves $C(n, r_1)$ choices. Decision 2 is which of the remaining $n - r_1$ spaces to fill with A_2's. Since there are r_2 of these characters, the second decision can be made in any one of $C(n - r_1, r_2)$ ways. Once the A_1's and A_2's have been placed, there are $n - r_1 - r_2$ positions remaining to be filled, and A_3's can be assigned to r_3 of them in $C(n - r_1 - r_2, r_3)$ ways, and so on. By the fundamental counting principle, the number of ways to make this sequence of decisions is the product

$$C(n, r_1) \times C(n - r_1, r_2) \times C(n - r_1 - r_2, r_3) \times \cdots \times C(n - r_1 - r_2 - \cdots - r_{k-1}, r_k).$$

(Because $r_1 + r_2 + \cdots + r_k = n$, the last factor in this product is $C(r_k, r_k) = 1$.) ■

It turns out that both binomial and multinomial coefficients have their unique qualities and uses. Keeping both is vastly more convenient than eliminating either.

Let's do some mathemagic. Pick a number, any number, just so long as it is an entry from Pascal's triangle. Suppose your pick happened to be $15 = C(6, 2)$. Starting with $C(2, 2)$, the first nonzero enry in column 2 (the third column of Fig. 1.5.1),

C(0,0)				
C(1,0)	C(1,1)			
C(2,0)	C(2,1)	**C(2,2)**		
		+		
C(3,0)	C(3,1)	**C(3,2)**	C(3,3)	
		+		
C(4,0)	C(4,1)	**C(4,2)**	C(4,3)	C(4,4)
		+		
C(5,0)	C(5,1)	**C(5,2)**	C(5,3)	C(5,4)
		+		
C(6,0)	C(6,1)	**C(6,2)**	C(6,3)	C(6,4)
C(7,0)	C(7,1)	C(7,2)	**C(7,3)**	C(7,4)

Figure 1.5.1

add the entries down to and including $C(6,2)$. The sum will be $C(7,3)$. Check it out:

$$\begin{aligned} C(2,2)+C(3,2)+C(4,2)+C(5,2)+C(6,2) &= 1+3+6+10+15 \\ &= 35 \\ &= C(7,3). \end{aligned}$$

The trick is an easy consequence of Pascal's relation and the fact that $C(2,2) = C(3,3)$. (See if you can reason it out before reading on.)

1.5.2 Chu's Theorem.* *If $n \geq r$, then*

$$\begin{aligned} \sum_{k=0}^{n} C(k,r) &= C(r,r)+C(r+1,r)+C(r+2,r)+\cdots+C(n,r) \\ &= C(n+1,r+1) \end{aligned}$$

(*where* $\sum_{k=0}^{n} C(k,r) = \sum_{k=r}^{n} C(k,r)$ *because* $C(k,r) = 0$, $k < r$).

Proof. Replace $C(r,r)$ with $C(r+1,r+1)$ and use Pascal's relation repeatedly to obtain

$$\begin{aligned} C(r+1,r+1)+C(r+1,r) &= C(r+2,r+1), \\ C(r+2,r+1)+C(r+2,r) &= C(r+3,r+1), \end{aligned}$$

and so on, ending with

$$C(n,r+1)+C(n,r) = C(n+1,r+1). \qquad \blacksquare$$

Chu's theorem has many interesting applications. To set the stage for one of them, we interrupt the mathematical discussion to relate a story about the young Carl Friedrich Gauss.† At the age of seven, Gauss entered St. Katharine's Volksschule in the duchy of Brunswick. One day his teacher, J. G. Büttner, assigned Gauss's class the problem of computing the sum

$$1+2+\cdots+100.$$

*Rediscovered many times, Theorem 1.5.2 can be found in Chu Shih-Chieh, *Precious Mirror of the Four Elements*, 1303.

†Gauss (1777–1855) is one of the half-dozen greatest mathematicians of the last millenium.

While his fellow pupils went right to work computing sums, Gauss merely stared at his slate and, after a few minutes, wrote

$$\frac{100 \times 101}{2} = 5050.$$

He seems to have reasoned that numbers can be added forwards or backwards,

$$\begin{aligned} 1 + \ \ 2 + \ \ 3 + \cdots + 98 + 99 + 100, \\ 100 + 99 + 98 + \cdots + \ \ 3 + \ \ 2 + \ \ \ 1, \end{aligned}$$

or even sidewards. Adding sidewards gives $1 + 100 = 101$, $2 + 99 = 101$, $3 + 98 = 101$, and so on. With each of the hundred columns adding to 101, the sum of the numbers in *both* rows, twice the total we're looking for, is 100×101.

Gauss's method can just as well be used to sum the first n positive integers:

$$\begin{aligned} 1 + 2 + \cdots + n &= \frac{n(n+1)}{2} \\ &= C(n+1, 2). \end{aligned} \tag{1.6}$$

Seeing the answer expressed as a binomial coefficient may seem a little contrived, but, with its left-hand side rewritten as $C(1,1) + C(2,1) + \cdots + C(n,1)$, Equation (1.6) is seen to be the $r = 1$ case of Chu's theorem!

There is a formula comparable to Equation (1.6) for the sum of the *squares* of the first n positive integers, namely,

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{1.7}$$

Once one has seen it (or guessed it), Equation (1.7) is easy enough to prove by induction. But, where did the formula come from in the first place? Chu's theorem! Summing both sides of

$$\begin{aligned} k^2 &= k + k(k-1) \\ &= C(k,1) + 2C(k,2), \end{aligned} \tag{1.8}$$

we obtain

$$\sum_{k=1}^{n} k^2 = \sum_{k=1}^{n} C(k,1) + 2\sum_{k=1}^{n} C(k,2).$$

Two applications of Chu's theorem (one with $r = 1$ and the other with $r = 2$) yield

$$\begin{aligned} 1^2 + 2^2 + \cdots + n^2 &= C(n+1,2) + 2C(n+1,3) \\ &= \frac{(n+1)n}{2} + 2\frac{(n+1)n(n-1)}{6} \\ &= n(n+1)\left[\frac{3+2(n-1)}{6}\right] \\ &= \frac{n(n+1)(2n+1)}{6}, \end{aligned}$$

precisely Equation (1.7).

What about summing mth powers? If we just had an analog of Equation (1.8), i.e., an identity of the form

$$k^m = \sum_{r=1}^{m} a_{r,m} C(k,r) \tag{1.9}$$

(where $a_{r,m}$ is independent of k, $1 \le r \le m$), we could sum both sides and use Chu's theorem to obtain

$$\begin{aligned} \sum_{k=1}^{n} k^m &= \sum_{k=1}^{n} \sum_{r=1}^{m} a_{r,m} C(k,r) \\ &= \sum_{r=1}^{m} a_{r,m} \sum_{k=1}^{n} C(k,r) \\ &= \sum_{r=1}^{m} a_{r,m} C(n+1, r+1). \end{aligned} \tag{1.10}$$

To see what's involved when $m = 3$, consider the equation

$$\begin{aligned} k^3 &= xC(k,1) + yC(k,2) + zC(k,3) \\ &= xk + \frac{1}{2}yk(k-1) + \frac{1}{6}zk(k-1)(k-2), \end{aligned}$$

which is equivalent to

$$6k^3 = (6x - 3y + 2z)k + (3y - 3z)k^2 + zk^3.$$

(Check it.) Equating coefficients of like powers of the integer variable k yields the system of linear equations

$$\begin{aligned} 6x - 3y + 2z &= 0, \\ 3y - 3z &= 0, \\ z &= 6, \end{aligned}$$

which has the unique solution $y = z = 6$ and $x = 1$. (Confirm this too.) Therefore,

$$k^3 = C(k,1) + 6C(k,2) + 6C(k,3) \tag{1.11}$$

or, in the language of Equation (1.9), $a_{1,3} = x = 1$, $a_{2,3} = y = 6$, and $a_{3,3} = z = 6$. Together, Equations (1.9)–(1.11) yield

$$\begin{aligned} 1^3 + 2^3 + \cdots + n^3 &= C(n+1,2) + 6C(n+1,3) + 6C(n+1,4) \\ &= \frac{n^2(n+1)^2}{4}. \end{aligned}$$

(Confirm *these* computations.)

Now we know where formulas for sums of powers of positive integers come from. They are consequences of Chu's theorem as manifested in Equations (1.9)-(1.10). From a theoretical point of view, that is all very well. The disagreeable part is the prospect of having to solve a system of m equations in m unknowns in order to identify the mystery coefficients $a_{r,m}$. In fact, there is an elegant solution to this difficulty!

In the form

$$\sum_{r=1}^{m} C(k,r)a_{r,m} = k^m,$$

Equation (1.9) is reminiscent of matrix multiplication. To illustrate this perspective, let $m = 6$ and consider that portion of Pascal's triangle lying in rows and columns numbered 1–6, i.e.,

$$\begin{array}{cccccc} 1 & & & & & \\ 2 & 1 & & & & \\ 3 & 3 & 1 & & & \\ 4 & 6 & 4 & 1 & & \\ 5 & 10 & 10 & 5 & 1 & \\ 6 & 15 & 20 & 15 & 6 & 1 \end{array}$$

Filling in the zeros corresponding to $C(n,r)$, $n < r \leq 6$, we obtain the matrix

$$C_6 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 & 0 \\ 4 & 6 & 4 & 1 & 0 & 0 \\ 5 & 10 & 10 & 5 & 1 & 0 \\ 6 & 15 & 20 & 15 & 6 & 1 \end{pmatrix}.$$

Anyone familiar with determinants will see that this matrix has an inverse. It is one of the most remarkable properties of binomial coefficients that C_n^{-1} can be obtained from C_n, just by sprinkling in some minus signs, e.g.,

$$C_6^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \\ 3 & -3 & 1 & 0 & 0 & 0 \\ -4 & 6 & -4 & 1 & 0 & 0 \\ 5 & -10 & 10 & -5 & 1 & 0 \\ -6 & 15 & -20 & 15 & -6 & 1 \end{pmatrix}.$$

(Before reading on, confirm that the product of these two matrices is the identity matrix, I_6.)

1.5.3 Definition. Let C_n be the $n \times n$ *Pascal matrix* whose (i,j)-entry is binomial coefficient $C(i,j)$, $1 \leq i, j \leq n$.

1.5.4 Alternating-Sign Theorem. *The Pascal matrix C_n is invertible; the (i,j)-entry of C_n^{-1} is $(-1)^{i+j}C(i,j)$.*

While it may seem a little like eating the dessert before the broccoli, let's defer the proof of the alternating-sign theorem to the end of the section and go directly to the application.

1.5.5 Theorem. *If m and r are positive integers, the coefficient of $C(k,r)$ in the equation $k^m = \sum_{r=1}^{m} a_{r,m}C(k,r)$ is given by*

$$a_{r,m} = \sum_{t=1}^{m} (-1)^{r+t} C(r,t) t^m.$$

This more-or-less explicit formula for $a_{r,m}$ eliminates the need to solve a system of equations. Put another way, Theorem 1.5.5 solves the corresponding system of m equations in m unknowns, once and for all, *for every* m.

Proof of Theorem 1.5.5. Suppose $n \geq m, r$. Let $A_n = (a_{i,j})$ be the $n \times n$ matrix of mystery coefficients (where $a_{r,m} = 0$ whenever $r > m$). Then, by Equation (1.9), the (k,m)-entry of C_nA_n is

$$\sum_{r=1}^{m} C(k,r)a_{r,m} = k^m,$$

$1 \leq k, m \leq n$. In other words, $C_nA_n = P_n$, where P_n is the $n \times n$ matrix whose (i,j)-entry is i^j. Thus, $A_n = C_n^{-1}P_n$, so the mystery coefficient $a_{r,m}$ is the (r,m)-entry of the matrix product $C_n^{-1}P_n$. ■

1.5.6 Example. Let's reconfirm Equation (1.11). By Theorem 1.5.5,

$$\begin{aligned} a_{1,3} &= (-1)^{1+1}C(1,1)1^3 = 1, \\ a_{2,3} &= (-1)^{2+1}C(2,1)1^3 + (-1)^{2+2}C(2,2)2^3 \\ &= -2 + 8 = 6, \\ a_{3,3} &= (-1)^{3+1}C(3,1)1^3 + (-1)^{3+2}C(3,2)2^3 + (-1)^{3+3}C(3,3)3^3 \\ &= 3 - 24 + 27 = 6; \end{aligned}$$

i.e., with $m = 3$, Equation (1.9) becomes $k^3 = C(k,1) + 6C(k,2) + 6C(k,3)$. □

In fact, it isn't necessary to compute $a_{r,m}$ for one value of r at a time, or even for one value of m at a time! Using matrices, we can calculate the numbers $a_{r,m}$, $1 \le r \le m$, $1 \le m \le n$, *all at once*!

1.5.7 Example. When $n = 4$,

$$P_4 = \begin{pmatrix} 1^1 & 1^2 & 1^3 & 1^4 \\ 2^1 & 2^2 & 2^3 & 2^4 \\ 3^1 & 3^2 & 3^3 & 3^4 \\ 4^1 & 4^2 & 4^3 & 4^4 \end{pmatrix}.$$

So,

$$\begin{aligned} C_4^{-1}P_4 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 3 & -3 & 1 & 0 \\ -4 & 6 & -4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 4 & 8 & 16 \\ 3 & 9 & 27 & 81 \\ 4 & 16 & 64 & 256 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 6 & 14 \\ 0 & 0 & 6 & 36 \\ 0 & 0 & 0 & 24 \end{pmatrix} = A_4. \end{aligned}$$

(Check the substitutions and confirm the matrix multiplication.) Observe that column 3 of A_4 recaptures Equation (1.11), column 2 reconfirms Equation (1.8), and column 1 reflects the fact that $k^1 = k = C(k,1)$. Column 4 is new:

$$k^4 = C(k,1) + 14C(k,2) + 36C(k,3) + 24C(k,4). \tag{1.12}$$

□

So much for the desert. It's time for the broccoli.

Proof of the Alternating-Sign Theorem. Given an $n \times n$ matrix $C = (c_{ij})$, recall that the $n \times n$ matrix $B = (b_{ij})$ is its inverse if and only if $CB = I_n$ if and only if $BC = I_n$. Let $C = C_n$ be the $n \times n$ Pascal matrix, so that $c_{ij} = C(i,j)$. In the context

of Theorem 1.5.4, we have a candidate for C^{-1}, namely, the matrix B, whose (i,j)-entry is $b_{ij} = (-1)^{i+j}C(i,j)$. With these choices, $CB = I_n$ if and only if

$$\sum_{k=1}^{n} C(i,k)(-1)^{k+j}C(k,j) = \delta_{i,j}, \tag{1.13a}$$

$1 \le i, j \le n$, and $BC = I_n$ if and only if

$$\sum_{k=1}^{n}(-1)^{i+k}C(i,k)C(k,j) = \delta_{i,j}, \tag{1.13b}$$

$1 \le i, j \le n$, where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise} \end{cases}$$

is the so-called *Knonecker delta*.

Let's prove Equation (1.13a). Because $C(i,k) = 0$, $k > i$, and $C(k,j) = 0$, $k < j$,

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

If $j > i$, the right-hand sum is empty, meaning that the left-hand sum is zero. (So far, so good.) If $i \ge k \ge j$, then (confirm it) $C(i,k)C(k,j) = C(i,j)C(i-j,k-j)$. Substituting this identity into the right-hand sum yields

$$\begin{aligned}\sum_{k=j}^{i}(-1)^{k+j}C(i,j)C(i-j,k-j) &= C(i,j)\sum_{k=j}^{i}(-1)^{j+k}C(i-j,k-j) \\ &= C(i,j)\sum_{r=0}^{i-j}(-1)^{r}C(i-j,r),\end{aligned}$$

where $r = k - j$. If $i = j$, this expression contains just one term, namely, $C(i,i)\times(-1)^0C(0,0) = 1$. So, to complete the proof of Theorem 1.5.4, it remains to establish the following. ■

1.5.8 Lemma. *If $n > 0$, then $\sum_{r=0}^{n}(-1)^r C(n,r) = 0$.*

1.5.9 Example. With $n = 5$, Lemma 1.5.8 becomes

$$C(5,0) - C(5,1) + C(5,2) - C(5,3) + C(5,4) - C(5,5) = 0,$$

which is an immediate consequence of symmetry: $C(5,2) = C(5,3)$, $C(5,1) = C(5,4)$, and $C(5,0) = C(5,5)$. If $n = 4$, the identity

$$\begin{aligned}C(4,0) - C(4,1) + C(4,2) - C(4,3) + C(4,4) &= 1 - 4 + 6 - 4 + 1 \\ &= 0,\end{aligned}$$

while just as valid, is a little less obvious. □

Proof of Lemma 1.5.8. The lemma follows from the binomial theorem, which will be taken up in section 1.7. It is easy enough, however, to give a direct proof. Observe that the conclusion is equivalent to

$$\sum_{r \text{ even}} C(n,r) = \sum_{r \text{ odd}} C(n,r),$$

i.e., the number of subsets of $T = \{1, 2, \ldots, n\}$ having even cardinality is equal to the number of subsets of T with odd cardinality.

Temporarily denote the family of all 2^n subsets of T by $\mathscr{F}$. We will prove the result by exhibiting a one-to-one, onto function[*] $f : \mathscr{F} \to \mathscr{F}$ such that $A \in \mathscr{F}$ has an even (odd) number of elements if and only if $f(A)$ has an odd (even) number. If $n = o(T)$ is odd, the function defined by $f(A) = T \backslash A = \{x \in T : x \notin A\}$, the complement of A, meets our needs. (This is the easy case, illustrated for $n = 5$ in Example 1.5.9.) If n is even, the function defined by

$$f(A) = \begin{cases} A \cup \{n\} & \text{when } n \notin A, \\ A \backslash \{n\} & \text{when } n \in A \end{cases}$$

satisfies our requirements. ■

1.5.10 Example. Some values of the function

$$f(A) = \begin{cases} A \cup \{4\} & \text{when } 4 \notin A, \\ A \backslash \{4\} & \text{when } 4 \in A \end{cases}$$

(corresponding to $n = 4$) are given in Fig. 1.5.2. □

A	$f(A)$
ϕ	{4}
{1}	{1,4}
{2}	{2,4}
{3,4}	{3}
{1,3,4}	{1,3}
{1,2,3,4}	{1,2,3}

Figure 1.5.2

1.5. EXERCISES

1 Prove that

(a) The sum $2 + 4 + 6 + \cdots + 2n$ of the first n even integers is $n(n + 1)$.

(b) The sum $1 + 3 + 5 + \cdots + (2n - 1)$ of the first n odd integers is n^2.

[*] One-to-one, onto functions are also known as *bijections*.

2 Evaluate

(a) $\sum_{i=1}^{n} i(i-1)$. **(b)** $\sum_{i=1}^{n} i(i+1)$.

(c) $\sum_{i=3}^{n}(2i-1)$. **(d)** $\sum_{i=1}^{n} i(i-1)(i-2)$.

3 A sequence of numbers $a_1, a_2, \ldots$ is *arithmetic* if there is a fixed constant c such that $a_{i+1} - a_i = c$ for all $i \geq 1$. For such a sequence, show that

(a) $a_{n+1} = a_1 + nc$. **(b)** $\sum_{i=1}^{n} a_i = \frac{1}{2}n(a_1 + a_n)$.

4 The proof of Theorem 1.5.1 given in the text is the *combinatorial proof*. Sketch the *algebraic proof*, i.e., write each of the binomial coefficients in terms of factorials and do lots of cancelling to obtain the multinomial coefficient.

5 Show that

(a) $C(r_k, r_k) \times C(r_{k-1} + r_k, r_{k-1}) \times \cdots \times C(r_1 + r_2 + \cdots + r_k, r_1)$
$$= \binom{n}{r_1, r_2, \ldots, r_k}.$$

(b) $\binom{r}{0} + \binom{r+1}{1} + \binom{r+2}{2} + \cdots + \binom{r+k}{k} = \binom{r+k+1}{k}$.

6 Use mathematical induction to prove that $1^3 + 2^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2$.

7 Confirm (by a brute-force computation) that

$$k^4 = C(k,1) + 14C(k,2) + 36C(k,3) + 24C(k,4).$$

8 Prove that $1^4 + 2^4 + \cdots + n^4 = \frac{1}{30}n(n+1)(2n+1)(3n^2+3n-1)$

(a) using Equations (1.9)–(1.10) and (1.12).

(b) using mathematical induction.

9 Solve for the coefficients $a_{r,5}, 1 \leq r \leq 5$, in the equation $k^5 = \sum_{r=1}^{5} a_{r,5}C(k,r)$

(a) using the matrix equation $A_5 = C_5^{-1}P_5$.

(b) by solving a system of five equations in five unknowns *without* using the matrix equation.

10 What is the formula for the sum of the fifth powers of the first n positive integers? (*Hint:* Lots of computations afford lots of opportunities to make mistakes. Confirm your formula for three or four values of n.)

11 Suppose f and g are functions of the positive integer variable n. If $f(n) = \sum_{r=1}^{n} C(n,r)g(r)$ for all $n \geq 1$, prove that $g(n) = \sum_{r=1}^{n}(-1)^{n+r}C(n,r)f(r)$ for all $n \geq 1$.

12 If $m \geq n$, prove that

(a) $\sum_{r=1}^{n} C(m,r)C(n-1,r-1) = C(m+n-1,n)$.

(b) $\sum_{r=1}^{n} rC(m,r)C(n,r) = nC(m+n-1,n)$.

13 Prove that $1 \times 2 + 2 \times 3 + 3 \times 4 + \cdots + n \times (n+1) = \frac{1}{3}n(n+1)(n+2)$.

14 Prove that $1 \times 2 \times 3 + 2 \times 3 \times 4 + \cdots + n(n+1)(n+2) = \frac{1}{4}n(n+1) \times (n+2)(n+3)$.

15 Prove Vandermonde's identity*: If m and n are positive integers, then

$$C(m,0)C(n,r) + C(m,1)C(n,r-1) + \cdots + C(m,r)C(n,0) = C(m+n,r).$$

16 Prove that $\sum_{r=0}^{n} C(n,r)^2 = C(2n,n)$. (Compare with Exercise 11, Section 1.2.)

17 How many of the $C(52,5)$ different five-card poker hands contain

(a) a full house? **(b)** four of a kind?

18 How many of the $C(52,13)$ different 13-card bridge hands contain

(a) all four aces? **(b)** a 4–3–3–3 suit distribution?

19 Show that

(a) $\sum_{r=1}^{n+1}(-1)^{r-1}[C(n,r-1)/r] = 1/(n+1)$.

(b) $\sum_{r=0}^{n}(-1)^{r}[C(n,r)/(r+1)] = 1/(n+1)$.

(c) $\sum_{r=1}^{n}(-1)^{r-1}[C(n,r)/r] = \sum_{k=1}^{n} 1/k$.

(d) $C_m^{-1}v^{\mathrm{t}} = w^{\mathrm{t}}$, where $v = (1/2, 1/3, \ldots, 1/[m+1])$ and $w = (1/2, -2/3, 3/4, -4/5, \ldots, [(-1)^{m+1}m/(m+1)])$.

(e) $C_m w^{\mathrm{t}} = v^{\mathrm{t}}$, where v and w are the vectors from part (d).

(f) Confirm the $m = 6$ case of part (e); i.e., write down the 6×6 matrix C_6 and confirm that $C_6 w^{\mathrm{t}} = v^{\mathrm{t}}$.

20 Let n be fixed. Denote the rth-power sum of the first $n-1$ positive integers by $g(r) = 1^r + 2^r + \cdots + (n-1)^r$. Show that

(a) $g(0) = n - 1$. **(b)** $g(1) = \frac{1}{2}n^2 - \frac{1}{2}n$.

(c) $g(2) = \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n$. **(d)** $g(3) = \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2$.

(e) $g(4) = \frac{1}{5}n^5 - \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n$.

21 The nth *Bernoulli number*, b_r, is the coefficient of n in the function $g(r)$ of Exercise 20. The first few Bernoulli numbers are exhibited in Fig. 1.5.3. Jakob Bernoulli (1654–1705) showed that the remaining coefficients in $g(r)$, $r \geq 1$,

r	0	1	2	3	4
b_r	1	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$

Figure 1.5.3. Bernoulli numbers.

* Named for Abnit-Theophile Vandermonde (1735–1796), who published the result in 1772 (469 years *after* it appeared in Chu Shih-Chieh's book).

can be expressed in terms of the b_r's by means of the identity

$$g(r) = \sum_{k=0}^{r} \frac{1}{k+1} C(r,k) b_{r-k} n^{k+1}.$$

(a) use the $r = 4$ case of this identity, along with Fig. 1.5.3, to recapture the expression for $g(4)$ in Exercise 20(e).

(b) Show that your solution to part (a) is consistent with Exercise 8.

(c) Compute $g(5)$.

(d) Show that your solution to part (c) is consistent with your solution to Exercise 10.

22 The Bernoulli numbers (Exercise 21) satisfy the implicit recurrence $\sum_{k=0}^{r} C(r+1,k) b_k = 0$, $r \geq 1$. Use this relation (and Fig. 1.5.3) to show that

(a) $b_5 = 0$. **(b)** $b_6 = \frac{1}{42}$. **(c)** $b_7 = 0$.

(d) $b_8 = -\frac{1}{30}$. **(e)** $b_9 = 0$. **(f)** $b_{10} = \frac{5}{66}$.

23 Let n be fixed. Prove that the function $g(r) = 1^r + 2^r + \cdots + (n-1)^r$, from Exercise 20, can expressed in the form $\sum_{k=1}^{r+1} c_{r,k} n^k$, where the coefficients satisfy the recurrence $(k+1)c_{r,k+1} = r c_{r-1,k}$ for all r, $k \geq 1$.

24 Use Exercises 20(e) and 23 and the fact that $g(r) = 1$ when $n = 2$ to compute $g(5)$.

25 Let r and s be integers, $0 \leq r < s$, and let

$$C_{[r,s]} = \begin{pmatrix} C(r,r) & C(r,r+1) & \cdots & C(r,s) \\ C(r+1,r) & C(r+1,r+1) & \cdots & C(r+1,s) \\ \vdots & \vdots & \ddots & \vdots \\ C(s,r) & C(s,r+1) & \cdots & C(s,s) \end{pmatrix}.$$

(a) Show that $C_{[1,n]} = C_n$.

(b) Exhibit $C_{[2,6]}$.

(c) Show that $C_{[r,s]}$ is an $(s-r+1)$-square matrix.

(d) Show that the (i,j)-entry of $C_{[r,s]}$ is $C(r+i-1, r+j-1)$.

(e) Show that $C_{[r,s]}$ is invertible.

(f) Exhibit $C_{[2,6]}^{-1}$.

(g) Prove that the (i,j)-entry of the inverse of $C_{[r,s]}$ is $(-1)^{i+j} C(r+i-1, r+j-1)$, $1 \leq i, j \leq s-r+1$.

(h) Let t be a nonnegative integer. If f and g are functions that satisfy $f(n) = \sum_{k=t}^{n} C(n,k) g(k)$ for all $n \geq t$, prove that $g(n) = \sum_{k=t}^{n} (-1)^{n+k} C(n,k) f(k)$ for all $n \geq t$.

26 The Fibonacci sequence (Exercise 19, Section 1.2) may be defined by $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$, $n \geq 1$.

(a) Show that $F_4 = F_2 + 2F_1 + F_0$.

(b) Show that $F_5 = F_3 + 2F_2 + F_1$.

(c) Show that $F_6 = F_3 + 3F_2 + 3F_1 + F_0$.

(d) Show that $F_7 = F_4 + 3F_3 + 3F_2 + F_1$.

(e) Given that $F_{2n+1} = \sum_{r=0}^{n} C(n,r)F_{r+1}$, prove that $F_{2n} = \sum_{r=0}^{n} C(n,r)F_r$.

(f) Prove that $F_n = \sum_{r=0}^{n} (-1)^{n+r} C(n,r)F_{2r}$. (*Hint:* Use part (e) and the $t = 1$ case of Exercise 25(h).)

27 If $C = C_{[0,m]}$ is the matrix from Exercise 25, show that $CK = L$, where

$$L = \begin{pmatrix} C(0,0) & C(1,1) & C(2,2) & C(3,3) & \cdots & C(m,m) \\ C(1,0) & C(2,1) & C(3,2) & C(4,3) & \cdots & C(m+1,m) \\ C(2,0) & C(3,1) & C(4,2) & C(5,3) & \cdots & C(m+2,m) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C(m,0) & C(m+1,1) & C(m+2,2) & C(m+3,3) & \cdots & C(m+m,m) \end{pmatrix},$$

$$K = \begin{pmatrix} C(0,0) & C(1,1) & C(2,2) & C(3,3) & \ldots & C(m,m) \\ 0 & C(1,0) & C(2,1) & C(3,2) & \cdots & C(m,m-1) \\ 0 & 0 & C(2,0) & C(3,1) & \cdots & C(m,m-2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & C(m,0) \end{pmatrix}.$$

28 For a fixed but arbitrary positive integer m, prove that the coefficients $a_{r,m}$, $1 \leq r \leq m$, in Equation (1.9) exist and are independent of k. (*Hint:* Show that any polynomial $f(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$ of degree at most m can be expressed (uniquely) as a linear combination of $p_0(x)$, $p_1(x), \ldots, p_m(x)$, where $p_0(x) = 1$ and $p_r(x) = (1/r!)x(x-1)\cdots(x-r+1)$, $r \geq 1$.)

1.6. FOUR WAYS TO CHOOSE

The prologues are over. ... It is time to choose.

— Wallace Stevens (*Asides on the Oboe*)

From its combinatorial definition, n-choose-r is the number of different r-element subsets of an n-element set. Because two subsets are equal if and only if they contain the same elements, $\binom{n}{r}$ depends on *what* elements are chosen, not when. In

computing $C(n, r)$, the *order* in which elements are chosen is irrelevant. The $C(5, 2) = 10$ two-element subsets of $\{\mathrm{L}, \mathrm{U}, \mathrm{C}, \mathrm{K}, \mathrm{Y}\}$ are

$$\{\mathrm{L}, \mathrm{U}\}, \{\mathrm{L}, \mathrm{C}\}, \{\mathrm{L}, \mathrm{K}\}, \{\mathrm{L}, \mathrm{Y}\}, \{\mathrm{U}, \mathrm{C}\}, \{\mathrm{U}, \mathrm{K}\}, \{\mathrm{U}, \mathrm{Y}\}, \{\mathrm{C}, \mathrm{K}\}, \{\mathrm{C}, \mathrm{Y}\}, \{\mathrm{K}, \mathrm{Y}\},$$

where, e.g., $\{\mathrm{L}, \mathrm{U}\} = \{\mathrm{U}, \mathrm{L}\}$. There are, of course, circumstances in which order is important.

1.6.1 Example. Consider all possible "words" that can be produced using two letters from the word LUCKY. By the fundamental counting principle, the number of such words is 5×4, twice $C(5, 2)$, reflecting the fact that order is important. The 20 possibilities are

$$\begin{aligned} &\mathrm{LU}, \mathrm{LC}, \mathrm{LK}, \mathrm{LY}, \mathrm{UC}, \mathrm{UK}, \mathrm{UY}, \mathrm{CK}, \mathrm{CY}, \mathrm{KY}, \\ &\mathrm{UL}, \mathrm{CL}, \mathrm{KL}, \mathrm{YL}, \mathrm{CU}, \mathrm{KU}, \mathrm{YU}, \mathrm{KC}, \mathrm{YC}, \mathrm{YK}. \end{aligned}$$ □

1.6.2 Definition. Denote by $P(n, r)$ the number of *ordered* selections of r elements chosen from an n-element set.

By the fundamental counting principle,

$$\begin{aligned} P(n, r) &= n(n-1)(n-2)\cdots(n-[r-1]) \\ &= n(n-1)(n-2)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!} \\ &= r!C(n, r). \end{aligned}$$

There is another way to arrive at this last identity: We may construe $P(n, r)$ as the number of ways to make a sequence of just two decisions. Decision 1 is which of the r elements to select, without regard to order, a decision having $C(n, r)$ choices. Decision 2 is how to order the r elements once they have been selected, and there are $r!$ ways to do that. By the fundamental counting principle, the number of ways to make the sequence of two decisions is $C(n, r) \times r! = P(n, r)$.

1.6.3 Example. Suppose nine members of the Alameda County School Boards Association meet to select a three-member delegation to represent the association at a statewide convention. There are $C(9, 3) = 84$ different ways to choose the delegation from those present. If the bylaws stipulate that each delegation be comprised of a delegate, a first alternate, and a second alternate, the nine members can comply from among themselves in any one of $P(9, 3) = 3!C(9, 3) = 504$ ways. □

1.6.4 Example. Door prizes are a common feature of fundraising luncheons. Suppose each of 100 patrons is given a numbered ticket, while its duplicate is placed in a bowl from which prize-winning numbers will be drawn. If the prizes are \$10, \$50, and \$150, then (assuming winning tickets are not returned to the

bowl) a total of $P(100, 3) = 970,200$ different outcomes are possible. If, on the other hand, the three prizes are each \$70, then the order in which the numbers are drawn is immaterial. In this case, the number of different outcomes is $C(100, 3) = 161,700$. □

Both $C(n, r)$ and $P(n, r)$ involve situations in which an object can be chosen at most once. We have been choosing *without replacement*. What about choosing *with* replacement? What if we recycle the objects, putting them back so they can be chosen again? How many ways are there to choose r things from n things with replacement? The answer depends on whether order matters. If it does, the answer is easy. The number of ways to make a sequence of r decisions each of which has n choices is n^r.

1.6.5 Example. How many different two-letter "words" can be produced using the "alphabet" $\{\mathrm{L}, \mathrm{U}, \mathrm{C}, \mathrm{K}, \mathrm{Y}\}$? If there are no restrictions on the number of times a letter can be used, then $5^2 = 25$ such words can be produced; i.e., there are 25 ways to choose 2 things from 5 with replacement if order matters. In addition to the 20 words from Example 1.6.1, there are five new ones, namely, LL, UU, CC, KK, and YY. □

This brings us to the fourth way to choose.

1.6.6 Example. In how many ways can $r = 10$ items be chosen from $\{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$ with replacement if order doesn't matter? As so often happens in combinatorics, the solution is most easily obtained by solving another problem that has the same answer. Suppose, e.g., A were chosen three times, B once, C twice, D not at all, and E four times. Associate with this selection the 14-letter "word"

$$|||-|-||--||||.$$

In this word, the "letter" | represents a tally mark. Since we are choosing 10 times, there are ten |'s. The dashes are used to separate tally marks corresponding to one letter from those that correspond to another. The first three |'s are for the three A's. The first dash separates the three A tallies from the single tally corresponding to the only B; the second dash separates the B tally from the two C tallies. There is no tally mark between the third and fourth dashes because there are no D's. Finally, the last four |'s represent the four E's. Since $\{A, B, C, D, E\}$ has $n = 5$ elements, we need 4 dashes to keep their respective tally marks separate. Conversely, any 14-letter word consisting of ten |'s and four —'s corresponds to a unique selection. The word $|||||||--|-||-$, e.g., correspons to seven A's, no B's, one C, two D's, and no E's.

Because the correspondence is one-to-one, the number of ways to select $r = 10$ things from $n = 5$ things with replacement where order doesn't matter is equal to the number of 14-letter words that can be made up from ten |'s and four —'s, i.e., to $C(14, 10) = 1001$. □

	Order matters	Order doesn't matter
Without replacement	$P(n,r)$	$C(n,r)$
With replacement	n^r	$C(r+n-1, r)$

Figure 1.6.1. The four ways to choose.

1.6.7 Theorem. *The number of different ways to choose r things from n things with replacement if order doesn't matter is* $C(r+n-1,r)$.

Proof. As in Example 1.6.6, there is a one-to-one correspondence between selections and $[r+(n-1)]$-letter words consisting of r tally marks and $n-1$ dashes. The number of such words is $C(r+n-1,r)$. ■

1.6.8 Example. Let's return to the door prizes of Example 1.6.4, but, this time, suppose that winning tickets are returned to the bowl so they have a chance to be drawn again. When the prizes are different, the $r=3$ winning tickets are chosen from the $n=100$ tickets in the bowl with replacement where order matters, and $100^3 = 1$ million different outcomes are possible. When the prizes are all the same (choosing with replacement when order doesn't matter), the number of different outcomes is only $C(3+100-1,3) = C(102,3) = 171,700$. □

The four ways to choose are summarized in Fig. 1.6.1. Because $C(r+n-1,r) = C(r+n-1,n-1) \neq C(r+n-1,n)$, it is important to remember that in the last column of the table each entry takes the form $C(*,r)$, where *r is the number of things chosen*, replacement or not. (Don't expect this second variable always to be labeled r.)

Choosing with replacement just means that elements may be chosen more than once. If order doesn't matter, then the only thing of interest is the multiplicity with which each element is chosen. As we saw in Example 1.6.6, $C(14,10) = 1001$ different outcomes are possible when choosing 10 times from $\{A,B,C,D,E\}$ with replacement when order doesn't matter. If, in one of these outcomes, A is chosen a times, B a total of b times, and so on, then

$$\mathrm{a}+\mathrm{b}+\mathrm{c}+\mathrm{d}+\mathrm{e}=10. \tag{1.14}$$

Evidently, each of the 1001 outcomes gives rise to a different nonnegative integer solution to Equation (1.14), and every nonnegative integer solution of this equation corresponds to a different outcome. In particular, Equation (1.14) must have precisely 1001 nonnegative integer solutions! The obvious generalization is this.

1.6.9 Corollary. *The equation* $x_1+x_2+\cdots+x_n=r$ *has exactly* $C(r+n-1,r)$ *nonnegative integer solutions.*

What about positive integer solutions? That's easy! The number of positive integer solutions to Equation (1.14) is equal to the number of nonnegative integer solutions to the equation

$$(a-1)+(b-1)+(c-1)+(d-1)+(e-1)=10-5,$$

namely, to $C(5+5-1,5)=C(9,5)=126$. [Of the 1001 nonnegative integer solutions to Equation (1.14), at least one variable is zero in all but 126 of them.]

1.6.10 Definition. A *composition*[*] of n having m *parts* is a solution, in positive integers, to the equation

$$n=x_1+x_2+\cdots+x_m. \tag{1.15}$$

Notice the change in notation. This is not deliberately meant to be confusing. Notation varies with context, and we are now moving on to a new idea. It might be useful to think of the integer variables n, r, k, m, etc., as a traveling company of players whose roles depend upon the demands of the current drama production.

A composition expresses n as a sum of parts; $7=5+2$ is a two-part composition of 7, not to be confused with $7=2+5$. In the first case, $x_1=5$ and $x_2=2$; in the second, $x_1=2$ and $x_2=5$. Never mind that addition is commutative. A composition is an *ordered* or *labeled* solution of Equation (1.15). The six two-part compositions of $n=7$ are $6+1$, $5+2$, $4+3$, $3+4$, $2+5$, and $1+6$, corresponding, e.g., to the six ways to roll a 7 with two dice (one red and one green).

1.6.11 Theorem. *The number of m-part compositions of n is* $C(n-1,m-1)$.

Proof. The number of positive integer solutions to Equation (1.15) is equal to the number of nonnegative integer solutions to

$$(x_1-1)+(x_2-1)+\cdots+(x_m-1)=n-m.$$

By Corollary 1.6.9, this equation has $C([n-m]+m-1,n-m)=C(n-1,n-m)=C(n-1,m-1)$ nonnegative integer solutions. ■

1.6.12 Example. The $C(6-1,3-1)=C(5,2)=10$ three-part compositions of 6 are illustrated in Fig. 1.6.2. □

1.6.13 Corollary. *The (total) number of compositions of n is* 2^{n-1}.

[*] The term was coined by Major Percy A. MacMahon (1854–1929). *Decomposition* might be a more descriptive word.

x_1	x_2	x_3
4	1	1
1	4	1
1	1	4
3	2	1
3	1	2
2	3	1
2	1	3
1	3	2
1	2	3
2	2	2

Figure 1.6.2

Proof. The number of compositions of n is the sum, as m goes from 1 to n, of the number of m-part compositions of n. According to Theorem 1.6.11, that sum is equal to

$$C(n-1,0) + C(n-1,1) + \cdots + C(n-1,n-1),$$

the sum of the numbers in row $n-1$ of Pascal's triangle. ■

By Corollary 1.6.13, there are $2^5 = 32$ different compositions of 6. Ten of them are tabulated in Fig. 1.6.2. You will be asked to list the remaining 22 compositions in Exercise 11, but why not do it now, while the idea is still fresh?

1.6.14 Example. How many integer solutions of $x + y + z = 20$ satisfy $x \geq 1$, $y \geq 2$, and $z \geq 3$? Solution: $x + y + z = 20$ if and only if $(x-1) + (y-2) + (z-3) = 14$. Setting $a = x - 1$, $b = y - 2$, and $c = z - 3$ transforms the problem into one involving the number of nonnegative integer solutions of $a + b + c = 14$. By Corollary 1.6.9, the answer is $C(14 + 3 - 1, 14) = 120$. □

1.6.15 Example. Some people are suspicious when consecutive integers occur among winning lottery numbers. This reaction is probably due to the common misconception that truly random numbers would be "spread out". Consider a simple example. Of the $C(6,3) = 20$ three-element subsets of $\{1,2,3,4,5,6\}$, how many fail to contain at least one pair of consecutive integers? Here is the complete list: $\{1,3,5\}$, $\{1,3,6\}$, $\{1,4,6\}$, and $\{2,4,6\}$.

What about the general case? Of the $C(n,r)$ r-element subsets of $S = \{1,2,\ldots,n\}$, how many do not contain even a single pair of consecutive integers? Recall the correspondence between r-element subsets of S and n-letter "words" consisting of r Y's and $n-r$ N's. In any such word, w, there will be some number, x_0, of N's that come before the first Y, some number x_1 of N's

between the first and second Y, some number x_2 of N's between the second and third Y, and so on, with some number x_r or N's coming after the last (rth) Y. Since w must contain a total of $n - r$ N's, it must be the case that

$$x_0 + x_1 + \cdots + x_r = n - r.$$

Every r-element subset of S corresponds to a unique solution of this equation, in nonnegative integers, and every nonnegative integer solution of this equation corresponds to a unique r-element subset of S. (Confirm that $C([n-r]+[r+1]-1, [n-r]) = C(n,r)$.)

In this correspondence between subsets and words, a subset contains no consecutive integers if and only if $x_i > 0$, $1 \le i \le r-1$. If we substitute $y_0 = x_0$, $y_r = x_r$, and $y_i = x_i - 1$, $1 \le i \le r-1$, then, as in Example 1.6.14, the answer to our problem is equal to the number of nonnegative integer solutions of

$$\begin{aligned} y_0 + y_1 + \cdots + y_r &= (n-r) - (r-1) \\ &= n - 2r + 1, \end{aligned}$$

i.e., to

$$\begin{aligned} C([n-2r+1]+[r+1]-1, [n-2r+1]) &= C(n-r+1, n-2r+1) \\ &= C(n-r+1, r). \end{aligned}$$

(Be careful, $C(n-r+1,r) \neq C(r+n-1,r)$.)

When $n = 6$ and $r = 3$, $C(6-3+1,3) = C(4,3) = 4$, confirming the result of the brute-force list in the first paragraph of this example. □

1.6. EXERCISES

1 Compute

(a) $P(5,3)$. **(b)** $C(5,3)$. **(c)** $C(5,2)$.

(d) $P(5,2)$. **(e)** $C(10,4)$. **(f)** $P(10,4)$.

(g) 7!.

2 Show that

(a) $nP(n-1,r) = P(n,r+1)$.

(b) $P(n+1,r) = rP(n,r-1) + P(n,r)$.

3 In how many ways can four elements be chosen from a seven-element set

(a) with replacement if order doesn't matter?

(b) without replacement if order does matter?

(c) without replacement if order doesn't matter?

(d) with replacement if order matters?

4 In how many ways can seven elements be chosen from a four-element set

(a) with replacement if order matters?

(b) with replacement if order doesn't matter?

(c) without replacement if order matters?

(d) without replacement if order doesn't matter?

5 In how many ways can four elements be chosen from a ten-element set

(a) with replacement if order matters?

(b) with replacement if order doesn't matter?

(c) without replacement if order doesn't matter?

(d) without replacement if order matters?

6 In how many ways can seven elements be chosen from a ten-element set

(a) without replacement if order matters?

(b) with replacement if order doesn't matter?

(c) without replacement if order doesn't matter?

(d) with replacement if order matters?

7 Show that multinomial coefficient $\binom{n}{n-r,1,1,\cdots,1} = P(n,r)$.

8 Compute the number of nonnegative integer solutions to

(a) $a+b=9$. **(b)** $a+b+c=9$.

(c) $a+b+c=30$. **(d)** $a+b+c+d=30$.

9 How many integer solutions of $a+b+c+d=30$ satisfy

(a) $d \geq 3$, $c \geq 2$, $b \geq 1$, $a \geq 0$?

(b) $a \geq 3$, $b \geq 2$, $c \geq 1$, $d \geq 0$?

(c) $a \geq 7$, $b \geq 2$, $c \geq 5$, $d \geq 6$?

(d) $a \geq -3$, $b \geq 20$, $c \geq 0$, $d \geq -2$?

10 Write down all 16 compositions of 5.

11 Ten of the 32 compositions of 6 appear in Fig. 1.6.2. Write down the remaining 22 compositions of 6.

12 How many compositions of 8 have

(a) 4 parts? **(b)** 4 or fewer parts?

(c) 6 parts? **(d)** 6 or fewer parts?

13 Prove that the inequality $x+y+z \leq 14$ has a total of 680 nonnegative integer solutions.

14 Prove that the inequality $x_1 + x_2 + \cdots + x_m \leq n$ has a total of $C(n+m, m)$ nonnegative integer solutions.

15 Starting with $F_0 = F_1 = 1$, the Fibonacci numbers satisfy the recurrence $F_n = F_{n-1} + F_{n-2}$, $n \geq 2$. Prove that

(a) $F_{k+n} = F_k F_n + F_{k-1} F_{n-1}$, $k, n \geq 1$.

(b) F_{2k+1} is a multiple of F_k, $k \geq 1$.

(c) F_{3k+2} is a multiple of F_k, $k \geq 1$.

16 Let F_n, $n \geq 0$, be the nth Fibonacci number. (See Exercise 15.) Prove that

(a) $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n+1} = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}, n \geq 1.$

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

(c) F_n and F_{n+1} are relatively prime.

17 Let n be a positive integer. Prove that there is a composition of n each of whose parts is a different Fibonacci number. (See Exercise 15.)

18 Let ρ_n be the number of compositions of n each of whose parts is greater than 1.

(a) Show that $\rho_6 = 5$ by writing down the compositions of 6 each of whose parts is at least 2.

(b) Show that $\rho_7 = 8$.

(c) If $n \geq 2$, prove that ρ_n is a Fibonacci number. (*Hint:* Exercise 19, Section 1.2.)

19 Let l_n be the number of compositions of n each of whose parts is at most 2. If $n \geq 1$, prove that $l_n = F_n$, the nth Fibonacci number.

20 The first "diagonal" of Pascal's triangle consists entirely of 1's. The second is comprised of the numbers 1, 2, 3, 4, 5, ... The fourth is illustrated in **boldface** in Fig. 1.6.3. Explain the relationship between the kth entry of the mth diagonal and choosing, with replacement, from $\{1, 2, \ldots, k\}$ where order doesn't matter.

1						
1	1					
1	2	1				
1	3	3	1			
1	**4**	6	4	1		
1	5	**10**	10	5	1	
1	6	15	**20**	15	6	1
			...			

Figure 1.6.3

21 Suppose five different door prizes are distributed among three patrons, Betty, Joan, and Marge. In how many different outcomes does

(a) Betty get three prizes while Joan and Marge each get one?

(b) Betty get one prize while Joan and Marge each get two?

22 Let A be the collection of all 32 compositions of 6. Let B be the 32-element family consisting of all subsets of $\{1, 2, 3, 4, 5\}$. Because $o(A) = o(B)$, there is a one-to-one correspondence between A and B.

(a) Prove that there are a total of 32! different one-to-one correspondences between A and B.

(b) Of the more than 2.6×10^{35} one-to-one correspondences between A and B, can any be described by an algorithm, or recipe, that transforms compositions into subsets?

23 What about choosing with *limited* replacement? Maybe the fundraising patrons in Examples 1.6.4 and 1.6.8 should be limited to at most two prizes. How many different outcomes are possible, under these terms of limited replacement, if there are 100 patrons and

(a) three different prizes? **(b)** three equal prizes?

(c) four different prizes? **(d)** four equal prizes?

24 Revisiting the "birthday paradox" (Exercises 20–21, Section 1.3), suppose each of k people independently chooses an integer between 1 and m (inclusive). Let p be the probability that some two of them choose the same number.

(a) Show that $p = 1 - P(m, k)/m^k$.

(b) M. Sayrafiezadeh showed that $p \doteq 1 - [1 - (k/2m)]^{k-1}$ as long as $k \leq m$, where "$\doteq$" means "about equal". Find the error in Sayrafiezadeh's estimate when $k = 23$ and $m = 365$.

25 Show that the number of compositions of n having k *or fewer* parts is $N(n-1, k-1) = C(n-1, 0) + C(n-1, 1) + \cdots + C(n-1, k-1)$ (a number involved in the sphere-packing bound of Section 1.4).

26 There is evidence in tomb paintings that ancient Egyptians used astragali (ankle bones of animals) to determine moves in simple board games. In later Greek and Roman times it was common to gamble on the outcome of throwing several astragali at once. When an astragalus is thrown, it can land in one of four ways. Compute the number of different outcomes when five astragali are thrown simultaneously.

27 Suppose you have four boxes, labeled A, B, C, and D. How many ways are there to distribute

(a) ten identical marbles among the four boxes?

(b) the numbers 0–9 among the four boxes?

28 Suppose, to win a share of the grand prize in the weekly lottery, you must match five numbers chosen at random from 1 to 49.

(a) Of the $C(49,5) = 1,906,884$ five-element subsets of $\{1, 2, \ldots, 49\}$, how many contain no consecutive integers? (*Hint*: Example 1.6.15.)

(b) Show that the probability of at least one pair of consecutive integers occurring in the weekly drawing is greater than $\frac{1}{3}$.

29 Prove that the (total!) number of subsets of $\{1, 2, \ldots, n\}$ that contain no two consecutive integers is F_{n+1}, the $(n+1)$st Fibonacci number. (See Exercises 15–19.)

1.7. THE BINOMIAL AND MULTINOMIAL THEOREMS

> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.
>
> — Robert Frost (*The Road Not Taken*)

Among the most widely known applications of binomial coefficients is the following.

1.7.1 Binomial Theorem. *If n is a nonnegative integer, then*

$$(x+y)^n = \sum_{r=0}^{n} C(n,r)x^r y^{n-r}.$$

Three applications of distributivity produce the identity

$$\begin{aligned}(x+y)^2 &= (x+y)(x+y)\\ &= x(x+y) + y(x+y)\\ &= xx + xy + yx + yy.\end{aligned} \tag{1.16}$$

The familiar next step would be to replace xx with x^2, $xy + yx$ with $2xy$, and so on, but let's freeze the action with Equation (1.16). As it stands, the right-hand side of this identity looks as if it could be a sum of two-letter "words". There is an alternative way to think about this word sum.

Starting with the expression $(x+y)(x+y)$, choose a letter, x or y, from the first set of parentheses, and one letter from the second set. Juxtapose the choices, in order, so as to produce what looks like a two-letter word. Do this in all possible ways, and sum the results. From this perspective, the right-hand side of

Equation (1.16) is a kind of *inventory*[*] of the four ways to make a sequence of two decisions. The term yx, e.g., records the sequence in which y is the choice for decision 1, namely, which letter to take from the first set of parentheses, and x is the choice for decision 2.

Applied to the expression

$$\begin{aligned}(x+y)^3 &= (x+y)^2(x+y)\\ &= (xx+xy+yx+yy)(x+y),\end{aligned}$$

this alternative view of distributivity suggests the following process: Select a two-letter word from $(xx+xy+yx+yy)$ and a letter from $(x+y)$. Juxtapose these selections (in order), so as to produce a three-letter word. Do this in all $(4\times 2=8)$ possible ways and sum, obtaining the following analog of Equation (1.16):

$$xxx+xyx+yxx+yyx+xxy+xyy+yxy+yyy. \tag{1.17}$$

A *variation* on this alternative view of distributivity would be to picture $(x+y)^3=(x+y)(x+y)(x+y)$ in terms, not of two decisions, but of three. Choose one of x or y from the first set of parentheses, one of x or y from the second set, and one of x or y from the third. String the three letters together (in order) to produce a three-letter word. Doing this in all $(2\times 2\times 2=8)$ possible ways and summing the results leads to Expression (1.17). However one arrives at that expression, replacing words like xyx with monomials like x^2y, and then combining like terms, produces the identity

$$(x+y)^3 = x^3+3x^2y+3xy^2+y^3. \tag{1.18}$$

The two variations on our alternative view of distributivity afford two different routes to a proof of the binomial theorem. One is inductive: Given the binomial *expansion* of $(x+y)^{n-1}$, the computation of $(x+y)^n$ is viewed in terms of two decisions, as in $(x+y)^n=(x+y)^{n-1}(x+y)$, and the proof is completed using Pascal's relation. In the second route, the expansion of $(x+y)^n$ is viewed in terms of n decisions.

Proof of Theorem 1.7.1. Taking the route "less traveled by", we evaluate the right-hand side of the equation

$$(x+y)^n=(x+y)(x+y)\cdots(x+y)$$

[*] Using distributivity to inventory the ways to make a sequence of decisions is an idea of fundamental importance in Pólya's enumeration theory (Chapter 3) and the theory of generating functions (Chapter 4).

in a series of steps. Begin by choosing one of x or y from the first set of parentheses, one from the second set, and so on, finally choosing one of x or y from the nth set. String the n choices together in order. Do this in all possible ways and sum the corresponding n-letter words. The resulting analog of expressions (1.16)–(1.17) is both an inventory of the 2^n ways to make a sequence of decisions and a vocabulary of all possible n-letter words that can be produced using the alphabet $\{x, y\}$. From this sum of words, the analog of Equation (1.18) is reached in two steps. Viewing x and y not as letters in an alphabet but as commuting variables, replace each n-letter word with a monomial of the form $x^r y^{n-r}$. Then combine like terms. In the resulting two-variable polynomial, the coefficient of $x^r y^{n-r}$ is the number of n-letter words in which r of the letters are x's and $n - r$ or them are y's. That number is known to us as $C(n, r)$. ■

Substituting $x = y = 1$ in the binomial theorem results in a new proof that

$$2^n = \sum_{r=0}^{n} C(n, r).$$

Setting $x = -1$ and $y = 1$ leads to another proof of Lemma 1.5.8, i.e.,

$$\sum_{r=0}^{n} (-1)^r C(n, r) = 0$$

for all $n \geq 1$. New results can be derived by making other substitutions, e.g., $x = 2$ and $y = 1$ yields an identity expressing 3^n in terms of powers of 2, namely,

$$3^n = \sum_{r=0}^{n} C(n, r) 2^r. \tag{1.19}$$

What happens if there are three variables? This is where the road less traveled by makes all the difference. Just as $(x + y)(x + y) \cdots (x + y)$ inventories the ways to make a sequence of n decisions each having two choices, $(x + y + z) \times (x + y + z) \cdots (x + y + z)$ inventories the ways to make a comparable sequence of decisions each having three choices. From this perspective, the process of expanding $(x_1 + x_2 + \cdots + x_k)^n$ is the same whether $k = 2$ or $k = 100$. Choose one of $x_1, x_2, \ldots, x_k$ from each of n sets of brackets. String the choices together, in order, obtaining an n-letter word. Do this in all k^n possible ways and sum. The resulting inventory is then simplified in two steps. First, each word is replaced with a monomial of (total) degree n, and then like terms are combined. At the end of this process, the coefficient of $x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}$, is the number of n-letter words that can be produced using r_1 copies of x_1, r_2 copies of $x_2, \ldots,$ and r_k copies of x_k. This proves the following generalization of the binomial theorem.

1.7.2 Multinomial Theorem. *If n is a nonnegative integer, then*

$$(x_1 + x_2 + \cdots + x_k)^n = \sum \binom{n}{r_1, r_2, \ldots, r_k} x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}, \tag{1.20}$$

where the sum is over all nonnegative integer solutions to the equation $r_1 + r_2 + \cdots + r_k = n$, and

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}.$$

Because some of the r's in Equation (1.20) may be zero, the sum is *not* over the k-part compositions of n. (Since $0! = 1$, the definition of multinomial coefficient is easily modified so as to permit zeros among its entries.)

1.7.3 Example. It isn't necessary to compute all $5^{10} = 9,765,625$ products in the expansion of $(a + b + c + d + e)^{10}$ just to determine the coefficient of $a^4 d^6$! From the multinomial theorem,

$$\binom{10}{4, 0, 0, 6, 0} = \frac{10!}{4!0!0!6!0!} = \frac{10!}{4!6!} = 210.$$

Observe that $210 = C(10, 4)$ is also the coefficient of $a^4 d^6$ in $(a + d)^{10}$, just as it should be. Setting $b = c = e = 0$ in $(a + b + c + d + e)^{10}$ has no effect on the coefficient of $a^4 d^6$. Also, observe that $\binom{10}{4,0,0,6,0} = \binom{10}{0,0,6,0,4}$. The coefficient of $c^6 e^4$ is also 210, reflecting the symmetry of $(a + b + c + d + e)^{10}$. We will return to this point momentarily. □

The usefulness of the multinomial theorem is not limited to picking off single coefficients. The expansion of all $3^4 = 81$ terms of $(x + y + z)^4$, e.g., looks like this:

$$\binom{4}{1, 2, 1} = 12 \qquad \binom{4}{1, 0, 3} = 4$$
$$x^4 + \cdots + \underset{\downarrow}{\text{——}} xy^2z + \cdots + \underset{\downarrow}{\text{——}} xz^3 + \cdots + z^4.$$

1.7.4 Example. What is the coefficient of xy in the expansion of $(1 + x + y)^5$? Solution: Because $xy = 1^3 xy$, the multinomial theorem can be applied directly. The answer is $\binom{5}{3,1,1} = 20$. Computing the coefficient of xy in $(2 + x + y)^5$ requires two steps. From the multinomial theorem, the coefficient of $2^3 xy$ is $\binom{5}{3,1,1} = 20$. So, the xy-term in the expansion of $(2 + x + y)^5$ is $20 \times 2^3 \times xy$, and the coefficient we're looking for is $20 \times 8 = 160$.

What about the coefficient of $x^3y^5z^2$ in $(2x-3y+4z)^{10}$? Since the coefficient of $(2x)^3(-3y)^5(4z)^2$ is $\binom{10}{3,5,2} = 2520$, the coefficient of $x^3y^5z^2$ must be $2520\times 2^3\times(-3)^5\times 4^2 = -78,382,080$. □

As with the binomial theorem, numerous identities can be obtained by making various substitutions for the variables in the multinomial theorem. Setting $x = y = z = 1$ in $(x+y+z)^n$, e.g., yields

$$3^n = \sum_{r+s+t=n} \binom{n}{r,s,t}. \tag{1.21}$$

Together with Equation (1.19), this produces the curious identity

$$\sum_{r=0}^{n} C(n,r)2^r = \sum_{r+s+t=n} \binom{n}{r,s,t}.$$

The multinomial theorem tells us that $x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ occurs among the k^n products in the expansion of $(x_1+x_2+\cdots+x_k)^n$ with multiplicity $\binom{n}{r_1,r_2,\ldots,r_k}$, but it does not tell us how many different monomial terms of the form $\binom{n}{r_1,r_2,\ldots,r_k}x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ occur in the expansion.

1.7.5 Theorem. *The number of different monomials of degree n in the k variables $x_1, x_2, \ldots, x_k$ is $C(n+k-1,n)$.*

Proof. From Corollary 1.6.9, the equation $r_1+r_2+\cdots+r_k = n$ has exactly $C(n+k-1,n)$ nonnegative integer solutions. ■

It makes perfect sense, of course, that the multinomial expansion of $(x_1+x_2+\cdots+x_k)^n$ should consist of $C(n+k-1,n)$ different monomial terms! In the first stage of computing

$$(x_1+x_2+\cdots+x_k)(x_1+x_2+\cdots+x_k)\cdots(x_1+x_2+\cdots+x_k),$$

each n-letter word identifies one of the k^n different ways to choose n times from $\{x_1,x_2,\ldots,x_k\}$ with replacement *where order matters*. After simplifying, each term in the resulting sum represents one of the $C(n+k-1,n)$ different ways to choose n times from $\{x_1,x_2,\ldots,x_k\}$ with replacement *where order doesn't matter.*

The multinomial expansion of $(x+y+z)^4$ is a *homogeneous* polynomial* comprised of $C(4+3-1,4)=15$ monomial terms, one of which is

$$\binom{4}{1,2,1}xy^2z = 12xy^2z.$$

Because $(x+y+z)^4$ is *symmetric*†, its multinomial expansion must be symmetric as well. Because switching x and y would interchange, e.g., xy^2z and x^2yz, these two monomials must have the same coefficient in the expansion of $(x+y+z)^4$. Indeed, $\binom{4}{1,2,1} = \binom{4}{2,1,1}$; the value of a multinomial coefficient does not change when two numbers in its bottom row are switched! Form either perspective, it is clear that

$$12x^2yz + 12xy^2z + 12xyz^2 = 12(x^2yz + xy^2z + xyz^2)$$

is a summand in the expansion of $(x+y+z)^4$, and it is natural to group these terms together. Organizing the remaining 12 terms in a similar fashion yields

$$\begin{aligned}(x+y+z)^4 = {} & (x^4+y^4+z^4) + 4(x^3y + x^3z + xy^3 + xz^3 + y^3z + yz^3) \\ & + 6(x^2y^2 + x^2z^2 + y^2z^2) + 12(x^2yz + xy^2z + xyz^2). \end{aligned} \tag{1.22}$$

The *minimal symmetric polynomials*‡ on the right-hand side of this equation have the symbolic names

$$\begin{aligned} M_{[4]}(x,y,z) &= x^4+y^4+z^4, \\ M_{[3,1]}(x,y,z) &= x^3y + x^3z + xy^3 + xz^3 + y^3z + yz^3, \\ M_{[2,2]}(x,y,z) &= x^2y^2 + x^2z^2 + y^2z^2, \\ M_{[2,1,1]}(x,y,z) &= x^2yz + xy^2z + xyz^2. \end{aligned}$$

Using this terminology, Equation (1.22) can be expressed as

$$\begin{aligned}(x+y+z)^4 = {} & M_{[4]}(x,y,z) + \binom{4}{3,1}M_{[3,1]}(x,y,z) + \binom{4}{2,2}M_{[2,2]}(x,y,z) \\ & + \binom{4}{2,1,1}M_{[2,1,1]}(x,y,z). \end{aligned} \tag{1.23}$$

*Each term has the same (total) degree, in this case four.

†Switching (any) two variables does not change the polynomial.

‡ "Minimal symmetric polynomial" is a descriptive name. these polynomials are known to experts as *monomial symmetric functions*.

1.7.6 Example. If $37y^3z$ is among the monomial terms of a symmetric polynomial $p(x, y, z)$, then

$$37(x^3y + x^3z + xy^3 + xz^3 + y^3z + yz^3) = 37M_{[3,1]}(x, y, z)$$

must be a summand of $p(x, y, z)$. □

There is nothing quite like a mountain of superscripts and subscripts to dull one's enthusiasm. So, there must be very good reasons for tolerating them in an introductory text. With a little getting used to, Equation (1.23) offers the best way to get a handle on the multinomial theorem, and a whole lot more! Let's see some more examples.

1.7.7 Example. By the multinomial theorem,

$$(x + y + z)^5 = \sum \binom{5}{a, b, c} x^a y^b z^c, \tag{1.24}$$

where the sum is over the nonnegative integer solutions to $a + b + c = 5$. The analog of Equation (1.23) is

$$\begin{aligned}(x + y + z)^5 = M_{[5]}(x, y, z) &+ \binom{5}{4, 1} M_{[4,1]}(x, y, z) + \binom{5}{3, 2} M_{[3,2]}(x, y, z) \\ &+ \binom{5}{3, 1, 1} M_{[3,1,1]}(x, y, z) + \binom{5}{2, 2, 1} M_{[2,2,1]}(x, y, z),\end{aligned} \tag{1.25}$$

where the $C(5 + 3 - 1, 5) = 21$ monomials of degree 5 have been organized into the minimal symmetric polynomials*

$$\begin{aligned}M_{[5]}(x, y, z) &= x^5 + y^5 + z^5, \\ M_{[4,1]}(x, y, z) &= x^4y + x^4z + xy^4 + xz^4 + y^4z + yz^4, \\ M_{[3,2]}(x, y, z) &= x^3y^2 + x^3z^2 + x^2y^3 + x^2z^3 + y^3z^2 + y^2z^3, \\ M_{[3,1,1]}(x, y, z) &= x^3yz + xy^3z + xyz^3, \\ M_{[2,2,1]}(x, y, z) &= x^2y^2z + x^2yz^2 + xy^2z^2.\end{aligned}$$ □

1.7.8 Example. The fifth power of a three-term sum was expanded in Example 1.7.7. Applying the multinomial theorem to the third power of a five-term sum produces

$$\begin{aligned}(a + b + c + d + e)^3 = M_{[3]}(a, b, c, d, e) &+ 3M_{[2,1]}(a, b, c, d, e) \\ &+ 6M_{[1,1,1]}(a, b, c, d, e),\end{aligned} \tag{1.26}$$

*It is just a coincidence that the 4*th* and 5*th* powers of $x + y + z$ involve four and five minimal symmetric polynomials, respectively. The 6*th* power involves seven.

where

$$\begin{aligned} M_{[3]}(a,b,c,d,e) &= a^3+b^3+c^3+d^3+e^3, \\ M_{[2,1]}(a,b,c,d,e) &= (a^2b+a^2c+a^2d+a^2e) \\ &\quad +(ab^2+b^2c+b^2d+b^2e)+\cdots+(ae^2+be^2+ce^2+de^2), \end{aligned} \tag{1.27}$$

and

$$\begin{aligned} M_{[1,1,1]}(a,b,c,d,e) &= abc+abd+abe+acd+ace+ade \\ &\quad +bcd+bce+bde+cde. \end{aligned} \tag{1.28}$$

□

1.7. EXERCISES

1 What is the coefficient of x^5 in the binomial expansion of

(a) $(x+y)^5$? **(b)** $(1+x)^7$? **(c)** $(1+x)^9$?

(d) $(2+x)^7$? **(e)** $(1+2x)^7$? **(f)** $(1-x)^9$?

(g) $(2-x)^4$? **(h)** $(2x+y)^4$? **(i)** $(2x-3y)^8$?

2 What is the coefficient of x^2y^3 in the multinomial expansion of

(a) $(x+y)^5$? **(b)** $(1+x+y)^5$?

(c) $(1+x+y)^8$? **(d)** $(2x-y)^5$?

(e) $(2+x+y)^5$? **(f)** $(3+2x-y)^8$?

(g) $(x-y+z)^5$? **(h)** $(-3+x-2y+z)^8$?

(i) $(1-2x+3y-4z)^7$? **(j)** $(1-2x+3y-4z)^4$?

3 Confirm Equation (1.21) in the case

(a) $n=4$ by setting $x=y=z=1$ in Equation (1.22).

(b) $n=5$ by setting $x=y=z=1$ in Equation (1.25).

4 Prove that $k^n = \sum \binom{n}{r_1,r_2,\ldots,r_k}$, where the sum is over all nonnegative integer sequences $(r_1,r_2,\ldots,r_k)$ that sum to n.

5 Consider the multinomial expansion of $(a+b+c+d+e)^3$ from Example 1.7.8.

(a) Explain why 3 and 6 are the correct coefficients of $M_{[2,1]}(a,b,c,d,e)$ and $M_{[1,1,1]}(a,b,c,d,e)$, respectively.

(b) Explain why $M_{[2,1]}(a,b,c,d,e)$ is a sum, not of $C(5,2) = 10$ monomials, but of $P(5,2) = 20$.

(c) Explain why $M_{[1,1,1]}(a,b,c,d,e)$ is a sum, not of $P(5,3) = 60$ monomials, but of $C(5,3) = 10$.

(d) Explain why the equation $5 + P(5,2) + C(5,3) = C(7,3)$ is a confirming instance of Theorem 1.7.5.

(e) Without doing any arithmetic, explain why $5 + 3P(5,2) + 6C(5,3) = 5^3$.

6 Prove the following special case of Exercise 10(c), Section 1.2, by differentiating $(1+x)^n$ and setting $x = 1$:

$$\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + r\binom{n}{r} + \cdots + n\binom{n}{n} = n2^{n-1}.$$

7 Of the 66 terms in the multinomial expansion of $(x+y+z)^{10}$, how many involve

(a) just one variable?

(b) exactly two variables?

(c) all three variables?

8 Show how Vandermonde's identity,

$$C(m,0)C(n,r) + C(m,1)C(n,r-1) + \cdots + C(m,r)C(n,0) = C(m+n,r),$$

follows from the equation $(x+1)^m(x+1)^n = (x+1)^{m+n}$.

9 Let n be a fixed but arbitrary positive integer. Multiply each multinomial coefficient of the form $\binom{n}{a,b,c,d}$ by $(-1)^{b+d}$ and add the results. Prove that the sum is zero.

10 Compute the coefficient of

(a) x^8 in $(x^2+1)^7$.

(b) x^8 in $(x^2+x)^7$.

(c) x^8 in $(x^2+x+1)^7$.

(d) x^5 in $(1+x+x^2)^7$.

(e) x^2y^2 in $(3+xy+xz+yz)^4$.

(f) $x^2y^2z^2$ in $(3+xy+xz+yz)^4$.

11 Let n be a positive integer and p a positive prime.

(a) Suppose $0 \le r_i < p$, $1 \le i \le n$. Prove that $\binom{p}{r_1,r_2,\ldots,r_n}$ is a multiple of p.

(b) Prove Fermat's "little theorem"*, i.e., that $n^p - n$ is an integer multiple of p.

*After Pierre de Fermat (1601–1665).

12 Give the (two-decision) inductive proof of the binomial theorem.

13 Write out all the terms of the minimal symmetric polynomial

(a) $M_{[6,4]}(x, y, z)$ **(b)** $M_{[5,5]}(x, y, z)$

14 Denote the coefficient of x^r in $(1 + x + x^2 + \cdots + x^{k-1})^n$ by $C_k(n, r)$.

(a) Show that $C_2(n, r) = C(n, r)$.

(b) Compute $C_3(3, 3)$.

(c) If $n > 1$, show that $C_k(n, r) = \sum_{i=0}^{k-1} C_k(n-1, r-i)$.

15 The multinomial expansion of $(x + y + z)^4$ can be expressed as a linear combination of four minimal symmetric polynomials and the expansion of $(x + y + z)^5$ as a linear combination of five. How many minimal symmetric polynomials are involved in the multinomial expansion of $(x + y + z)^{10}$? (Two of them appear in Exercise 13.)

16 It follows from Theorem 1.6.11 that the number of compositions of n having k or fewer parts is $N(n-1, k-1) = C(n-1, 0) + C(n-1, 1) + \cdots + C(n-1, k-1)$. By Theorem 1.7.5, there are $C([n-1]+k, k-1)$ different monomials in the multinomial expansion of $(x_1 + x_2 + \cdots + x_k)^n$. It does not seem to follow, however, that $N(n-1, k-1) = C([n-1]+k, k-1)$. With $n = 6$ and $k = 3$, $N(5, 2) = 16$ while $C([6-1]+3, 3-1) = C(8, 2) = 28$. Write out enough terms in the expansion of $(x + y + z)^6$ to explain where the numbers 16 and 28 come from.

17 Use Theorem 1.5.1 and the binomial theorem to give another proof of the multinomial theorem.

18 Exercise 14, Section 1.1, asks for an explicit listing of the 24 (exact) positive integer divisors of $360 = 2^3 3^2 5$. *Without* doing any arithmetic, explain why the *sum* of these 24 divisors is given by the product $(1 + 2 + 2^2 + 2^3) \times (1 + 3 + 3^2)(1 + 5)$.

19 Suppose the prime factorization of $n = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$. Prove that the sum of the divisors of n is the product

$$\prod_{t=1}^{k} \left(\sum_{s=0}^{r_t} p_t^s \right).$$

20 Explain how the binomial theorem can be used to prove that $\sum_{r=0}^{n} P(r) = 1$, where $P(r) = C(n, r) p^r q^{n-r}$ is the binomial probability distribution of Section 1.3, Equation (1.5).

21 For a fixed but arbitrary integer $n \geq 2$, define $g(r) = M_{[r]}(1, 2, \ldots, n-1) = 1^r + 2^r + \cdots + (n-1)^r$.

(a) Prove that $\sum_{r=0}^{k} C(k+1, r) g(r) = n^{k+1} - 1$.

(b) Given $g(0), g(1), \ldots, g(r)$, the equation in part (a) can be used to solve for $g(r+1)$. Starting from $g(0) = n - 1$, use this method to compute $g(1)$, $g(2)$, $g(3)$, and $g(4)$.

(c) Compare and contrast with the approach suggested by Section 1.5, Exercise 11.

(d) Explain the connection with Bernoulli numbers (Section 1.5, Exercises 20–22).

22 Show that $\sum_{r=0}^{25} C(50, r)C(50 - r, 25 - r) = 2^{25}C(50, 25)$.

23 Compute

(a) $\sum_{r=25}^{50} C(50, r)C(r, 25)$.

(b) $\sum_{r=0}^{25} (-1)^r C(50, r)C(50 - r, 25)$.

24 Prove that the alternative view of distributivity used to prove the binomial and multinomial theorems is valid, i.e., suppose $S_1, S_2, \ldots, S_n$ are sums of algebraic terms. Prove that $S_1 \times S_2 \times \cdots \times S_n$ is the sum of all products that can be obtained by choosing one term from each sum, multiplying the choices together, doing this in all $o(S_1) \times o(S_2) \times \cdots \times o(S_n)$ possible ways, and adding the resulting products. (*Hint:* Induction on n.)

1.8. PARTITIONS

Something there is that doesn't love a wall.

— Robert Frost (*Mending Wall*)

In the last section, we grouped the $C(n + k - 1, n)$ different monomials from the multinomial expansion of $(x_1 + x_2 + \cdots + x_k)^n$ into certain minimal symmetric polynomials with symbolic names like $M_{[4,1]}$ and $M_{[2,2,1]}$.

1.8.1 Definition. A *partition* of n having *m parts* is an unordered collection of m positive integers that sum to n.

1.8.2 Example. The number 6 is said to be *perfect*[*] because it is the sum of its proper divisors: $6 = 1 + 2 + 3$. Since addition is commutative, this sum could just as well have been written $2 + 3 + 1$. In this context, $1 + 2 + 3$ is the same as

[*]A Christian theologian once argued that God, who could have created the universe in an instant, chose instead to labor for 6 days in order to emphasize the perfection of His creation. (It is just an accident that this book has 6 chapters.)

$2 + 3 + 1$ but different from $4 + 2$. In expressing the prefection of 6, what interests us is the unordered collection of its proper divisors, the partition whose parts are 3, 2, and 1. □

Two partitions of n are equal if and only if they have the same parts with the same multiplicities. By way of contrast, a *composition* of n (Definition 1.6.10) is an *ordered* collection of positive integers that sum to n. Compositions are sometimes called *ordered partitions*. Two compositions are equal if and only if they have the same parts with the same multiplicities, *in the same order.*

Our discussion of partitions will be simplified by the adoption of some notation.

1.8.3 Definition. An m-part partition of n is represented by a sequence $\pi = [\pi_1, \pi_2, \ldots, \pi_m]$ in which the parts are arranged so that $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_m > 0$. The number of parts is the *length* of π, denoted $\ell(\pi) = m$. The shorthand expression $\pi \vdash n$ signifies that "π is a partition of n".

In ordinary English usage, arranging the parts of a partition from largest to smallest would typically be called "orderning" the parts. This semantic difficulty is the source of more than a little confusion. It is precisely because a partition is unordered that we are free to arrange its parts any way we like. The 5 cards comprising a poker hand can be arranged in any one of $5! = 120$ different ways. But, no matter how the cards are arranged or rearranged, the poker hand is the same. So it is with partitions. A composition, on the other hand, is some specified arrangement of the parts of a partition. By convention (Definition 1.8.3), we uniformly choose one such composition to represent each partition.

1.8.4 Example. The three-part partitions of 6 are [4, 1, 1], [3, 2, 1], and [2, 2, 2]. There are 3 ways to arrange the parts of [4, 1, 1], 6 ways to arrange the parts of [3, 2, 1], but only one way to arrange the parts of [2, 2, 2]. Taken together, these 10 arrangements comprise the compositions of 6 having 3 parts (as illustrated in Fig. 1.6.2). □

Already it seems convenient to introduce some additional shorthand notation. Rather than [4, 1, 1] and [2, 2, 2], we will write $[4, 1^2]$ and $[2^3]$, respectively. Similarly, the partition [5, 5, 3, 3, 3, 3, 2, 2, 2, 1] is abbreviated $[5^2, 3^4, 2^3, 1]$. In this notation superscripts denote, not exponents, but multiplicities. In the 10-part partition $[5^2, 3^4, 2^3, 1]$, the piece 3^4 contributes, not $3 \times 3 \times 3 \times 3 = 81$, but $3 + 3 + 3 + 3 = 12$ to the sum

$$5 + 5 + 3 + 3 + 3 + 3 + 2 + 2 + 2 + 1 = 29.$$

The m-part compositions of n were counted in Theorem 1.6.11. (They number $C(n-1, m-1)$.) Counting the m-part partitions of n is not so easy. Let's begin by giving this number a name.

n \ m	1	2	3	4	5	6	7
1	1						
2	1	1					
3	1	1	1				
4	1	$p_2(4)$	1	1			
5	1	2	2	1	1		
6	1	$p_2(6)$	3	$p_4(6)$	1	1	
7	1	$p_2(7)$	$p_3(7)$	$p_4(7)$	$p_5(7)$	1	1
				...			

Figure 1.8.1. The partition triangle.

1.8.5 Definition. The number of m-part partitions of n is denoted $p_m(n)$.

1.8.6 Example. From Example 1.8.4, $p_3(6) = 3$. The seven partitions of 5 are $[5]$, $[4,1]$, $[3,2]$, $[3,1^2] = [3,1,1]$, $[2^2,1] = [2,2,1]$, $[2,1^3] = [2,1,1,1]$, and $[1^5] = [1,1,1,1,1]$, having lengths 1, 2, 2, 3, 3, 4, and 5, respectively. Hence, $p_1(5) = 1$, $p_2(5) = 2$, $p_3(5) = 2$, $p_4(5) = 1$, and $p_5(5) = 1$. □

Because $[n]$ is the only partition of n having just one part and, at the other extreme, $[1^n]$ is the only partition of n having n parts, $p_1(n) = 1 = p_n(n)$ for all n. If $n \geq 2$, then $[2,1^{n-2}]$ is the only partition of n having length $n-1$, so $p_{n-1}(n) = 1$ as well.

The numbers $p_m(n)$ are displayed in the Pascal-like *partition triangle* of Fig. 1.8.1, where it is understood that $p_m(n) = 0$ when $m > n$. What is needed is a Pascal-like relation that would allow the entries of this triangle to be filled in a row at a time.

1.8.7 Theorem. *The number of m-part partitions of n is* $p_m(n) = p_{m-1}(n-1) + p_m(n-m)$, $1 < m < n$.

Proof. If π is an m-part partition of n, then $\pi_m = 1$ or it doesn't. There are $p_{m-1}(n-1)$ partitions of the first kind. Because $\pi \leftrightarrow [\pi_1 - 1, \pi_2 - 1, \ldots, \pi_m - 1]$ is a one-to-one correspondence between the m-part partitions of n satisfying $\pi_m > 1$ and the m-part partitions of $n-m$, there must be $p_m(n-m)$ partitions of the second kind. ■

From Theorem 1.8.7, $p_2(4) = p_1(3) + p_2(4-2) = p_1(3) + p_2(2) = 1 + 1 = 2$. (The two-part partitions of 4 are [3, 1] and $[2^2]$.) Similarly, $p_2(6) = p_1(6) + p_2(4) = 1 + 2 = 3$, and $p_4(6) = p_3(5) + p_4(2) = 2 + 0 = 2$. This completes Fig. 1.8.1 through row 6. Rows 7–10 are completed in Fig. 1.8.2.

1.8.8 Definition. Denote the number of partitions of n by $p(n) = p_1(n) + p_2(n) + \cdots + p_n(n)$.

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

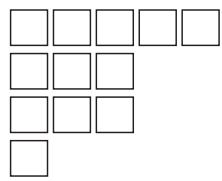
Figure 1.8.2. The partition numbers $p_m(n)$.

Just as the nth row sum of Pascal's triangle is 2^n, the total number of subsets of an n-element set, the nth row sum of the partition triangle is $p(n)$, the total number of partitions of n. Summing, rows 9 and 10 of Fig. 1.8.2, e.g., yields the partition numbers $p(9) = 30$ and $p(10) = 42$.*

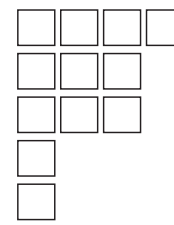
If π is an m-part partition of n, its *Ferrers diagram*,† $F(\pi)$, consists of n "boxes" arrayed in m left-justified rows, where the number of boxes in row i is π_i. The diagrams for $[5, 3^2, 1]$ and $[4, 3^2, 1^2]$, e.g., appear in Fig. 1.8.3.

1.8.9 Definition. The *conjugate* of $\pi \vdash n$ is the partition $\pi^* \vdash n$ whose jth part is the number of boxes in the jth column of $F(\pi)$.

Because the number of boxes in row j of $F(\pi^*)$ is equal to the number of boxes in column j of $F(\pi)$ for all j, the two diagrams are *transposes* of each other. In



$F([5, 3^2, 1])$



$F([4, 3^2, 1^2])$

Figure 1.8.3. Two Ferrers diagrams.

*The partition numbers grow rapidly with n. MacMahon showed, e.g., that $p(200) = 3,972,999,029,388$.

†Named for Norman Macleod Ferrers (1829–1903) but possibly used earlier by J. J. Sylvester (1814–1897).

particular, partition $\alpha = \pi^*$ if and only if $\alpha^* = \pi$. This situation is illustrated in Fig. 1.8.3 for the conjugate pair $[5, 3^2, 1]$ and $[4, 3^2, 1^2]$.

The number of boxes in the jth column of $F(\pi)$ is equal to the number of rows of $F(\pi)$ that contain at least j boxes, i.e., π_j^* is equal to the number of parts of π that are not less than j. Said another way, the jth part of π^* is

$$\pi_j^* = o(\{i : \pi_i \geq j\}). \tag{1.29}$$

1.8.10 Theorem. *The number of m-part partitions of n is equal to the number of partitions of n whose largest part is m.*

Proof. If π is an m-part partition of n, then m is the number of boxes in the first column of $F(\pi)$, i.e., $m = \pi_1^*$, the largest part of π^*. Hence, in the one-to-one correspondence between partitions and their conjugates, the set of m-part partitions corresponds to the set of partitions whose largest part is m. ■

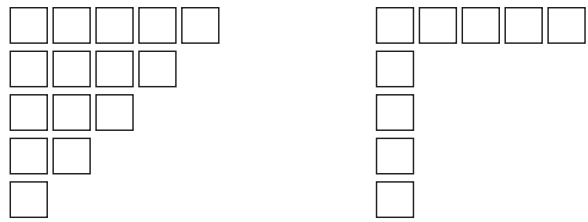
1.8.11 Definition. Partition π is *self-conjugate* if $\pi^* = \pi$.

1.8.12 Example. Because $\pi = \pi^*$ if and only if $F(\pi) = F(\pi^*) = F(\pi)^t$, the transpose of $F(\pi)$, π is self-conjugate if and only if its Ferrers diagram is symmetric about the "main diagonal". Thus, merely by glancing at Fig. 1.8.4, one sees that $[5, 4, 3, 2, 1]$ and $[5, 1^4]$ are self-conjugate partitions. On the other hand, without a Ferrers diagram to look at, it is much less obvious that $[5^2, 4, 3, 2]$ is self-conjugate. □

Knowing something about partitions, we can now give a formal definition of "minimal symmetric polynomial".

1.8.13 Definition. Let k and n be positive integers. Suppose π is an m-part partition of n. If $k \geq m$, the *minimal symmetric polynomial*

$$M_\pi(x_1, x_2, \ldots, x_k) = \sum x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k},$$



F ([5, 4, 3, 2, 1]) F ([5, 1^4])

Figure 1.8.4. Two self-conjugate partitions.

where the sum is over all different rearrangements $(r_1, r_2, \ldots, r_k)$ of the k-tuple $(\pi_1, \pi_2, \ldots, \pi_m, 0, \ldots, 0)$ that is obtained by appending $k - m$ zeros to the end of π. If $k < m$, then $M_\pi(x_1, x_2, \ldots, x_k) = 0$.

If, e.g., $\pi = [\pi_1, \pi_2] = [2, 2]$ and $k = 3$, the *different rearrangements* of $(\pi_1, \pi_2, 0)$ are (2, 2, 0), (2, 0, 2), and (0, 2, 2), and *not* the six different-*looking* ways to rearrange the symbols π_1, π_2, and 0. In particular,

$$M_{[2,2]}(x, y, z) = x^2y^2 + x^2z^2 + y^2z^2.$$

If $\pi \vdash n$ and $m = \ell(\pi) \leq k$, then each monomial $x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ in Definition 1.8.13 has (total) degree $r_1 + r_2 + \cdots + r_k = \pi_1 + \pi_2 + \cdots + \pi_m = n$, i.e., $M_\pi(x_1, x_2, \ldots, x_k)$ is homogeneous of degree n.

1.8.14 Example. From Fig. 1.8.2, there are $p_1(6) + p_2(6) + p_3(6) = 1 + 3 + 3 = 7$ different partitions of 6 having at most three parts. Hence, there are 7 different minimal symmetric polynomials of degree 6 in the variables x, y, and z, namely,

$$\begin{aligned}
M_{[6]}(x, y, z) &= x^6 + y^6 + z^6,\\
M_{[5,1]}(x, y, z) &= x^5y + x^5z + xy^5 + xz^5 + y^5z + yz^5,\\
M_{[4,2]}(x, y, z) &= x^4y^2 + x^4z^2 + x^2y^4 + x^2z^4 + y^4z^2 + y^2z^4,\\
M_{[3^2]}(x, y, z) &= x^3y^3 + x^3z^3 + y^3z^3,\\
M_{[4,1^2]}(x, y, z) &= x^4yz + xy^4z + xyz^4,\\
M_{[3,2,1]}(x, y, z) &= x^3y^2z + x^3yz^2 + x^2y^3z + x^2yz^3 + xy^3z^2 + xy^2z^3,
\end{aligned}$$

and

$$M_{[2^3]}(x, y, z) = x^2y^2z^2. \qquad \square$$

Minimal symmetric polynomials are to symmetric polynomials what atoms are to molecules. they are the basic building blocks.

1.8.15 Theorem. *The polynomial $f = f(x_1, x_2, \ldots, x_k)$ is symmetric in $x_1, x_2, \ldots, x_k$ if and only if it is a linear combination of minimal symmetric polynomials.*

Proof. Because minimal symmetric polynomials are symmetric, any linear combination of minimal symmetric polynomials in $x_1, x_2, \ldots, x_k$ is symmetric.

Conversely, suppose $cx_1^{s_1}x_2^{s_2}\cdots x_k^{s_k}$ is among the nonzero terms of $f(x_1, x_2, \ldots, x_k)$. Then $(s_1, s_2, \ldots, s_k)$ is a rearrangement of $(\alpha_1, \alpha_2, \ldots, \alpha_m, 0, \ldots, 0)$ for some partition α. Because f is symmetric, $cx_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ must occur among its terms for *every* rearrangement $(r_1, r_2, \ldots, r_k)$ of $(\alpha_1, \alpha_2, \ldots, \alpha_m, 0, \ldots, 0)$, i.e.,

$cM_\alpha(x_1, x_2, \ldots, x_k)$ is a summand of f. Therefore, $f(x_1, x_2, \ldots, x_k) - cM_\alpha(x_1, x_2, \ldots, x_k)$ is a symmetric polynomial with fewer terms than f, and the result follows by induction. ■

1.8.16 Example. Let

$$f(a,b,c,d) = 2a^3 - a^2b - a^2c - a^2d - ab^2 + abc + abd - ac^2 + acd - ad^2 + 2b^3 - b^2c - b^2d - bc^2 + bcd - bd^2 + 2c^3 - c^2d - cd^2 + 2d^3.$$

Probably the easiest way to confirm that this polynomial is symmetric is to express it as

$$f(a,b,c,d) = 2M_{[3]}(a,b,c,d) - M_{[2,1]}(a,b,c,d) + M_{[1^3]}(a,b,c,d). \qquad \square$$

There are, of course, easier ways to verify that the polynomial $f(x_1, x_2, \ldots, x_k) = (x_1 + x_2 + \cdots + x_k)^n$ is symmetric than by expressing it as a linear combination of minimal symmetric polynomials. On the other hand, because it *is* symmetric, $f(x_1, x_2, \ldots, x_k)$ *is* a linear combination of minimal symmetric polynomials. What combination? The answer to that question is what the multinomial theorem is all about:

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\pi \vdash n} \binom{n}{\pi} M_\pi(x_1, x_2, \ldots, x_k), \tag{1.30}$$

where the coefficient $\binom{n}{\pi}$ is an abbreviation for the multinomial coefficient whose bottom row consists of the $\ell(\pi)$ parts of π. (Recall that $M_\pi(x_1, x_2, \ldots, x_k) = 0$ whenever $k < \ell(\pi)$.)

1.8.17 Example. Together with Example 1.8.14, Equation (1.30) yields

$$\begin{aligned}(x+y+z)^6 = {} & M_{[6]}(x,y,z) + 6M_{[5,1]}(x,y,z) + 15M_{[4,2]}(x,y,z) \\ & + 20M_{[3^2]}(x,y,z) + 30M_{[4,1^2]}(x,y,z) \\ & + 60M_{[3,2,1]}(x,y,z) + 90M_{[2^3]}(x,y,z). \end{aligned} \qquad \square$$

1.8. EXERCISES

1 Explicitly write down

(a) all 11 partitions of 6.

(b) all 8 partitions of 7 having at most three parts.

(c) all 8 partitions of 7 whose largest part is at most three.

2 Show that

(a) $p_{n-2}(n) = 2, \quad n \geq 4.$

(b) $p_{n-3}(n) = 3, \quad n \geq 6.$

(c) for all $n \geq 6$, the last four (nonzero) numbers in row n of the partition triangle are 3, 2, 1, 1.

(d) $p_2(n) = \lfloor n/2 \rfloor$, the greatest integer not exceeding $\frac{1}{2}n$.

3 Compute rows 11–15 of the partition triangle.

4 Evaluate

(a) $p(11)$. **(b)** $p(12)$.

(c) $p(13)$. **(d)** $p(14)$.

5 The number of partitions of n into three or fewer parts turns out to be the nearest integer to $\frac{1}{12}(n+3)^2$.

(a) Confirm this fact for $1 \leq n \leq 6$.

(b) Confirm this fact for $7 \leq n \leq 10$.

(c) Determine the number of different minimal symmetric polynomials, in three variables, of degee $n = 27$.

6 How many different eight-part compositions can be produced by rearranging the parts of the partition

(a) $[5^3, 4, 2^4]$? **(b)** $[2^5, 1^3]$?

(c) $[8, 7, 6, 5, 4, 3, 2, 1]$?

(*Hint*: Don't try to write them all down.)

7 Confirm, by writing them all down, that there are $p_3(9)$ four-part partitions $\pi \vdash 10$ that satisfy $\pi_4 = 1$.

8 Confirm Theorem 1.8.10 for the pair

(a) $n = 5$ and $m = 2$. **(b)** $n = 6$ and $m = 3$.

(c) $n = 10$ and $m = 3$. **(d)** $n = 10$ and $m = 5$.

9 Prove that the partition number $p(n) \geq 2^{\lfloor \sqrt{n} \rfloor}$ for all sufficiently large n.

10 Exhibit Ferrers diagrams for all the self-conjugate partitions of

(a) 6. **(b)** 10. **(c)** 17.

11 Let $p_{\text{odd}}(n)$ be the number of partitions of n each of whose parts is odd and $p_{\text{dist}}(n)$ be the number of partitions of n having distinct parts. It is proved in Section 4.3 that $p_{\text{odd}}(n) = p_{\text{dist}}(n)$ for all n. Confirm this result now for the case

(a) $n = 5$. **(b)** $n = 6$. **(c)** $n = 7$. **(d)** $n = 8$.

12 The first odd "abundant" number is 945.

(a) How many positive integer divisors does 945 have?

(b) Sum up the "proper" divisors of 945 (those divisors less than 945).

(c) What do you suppose an "abundant" number is?

13 Prove that the number of partitions of n with at most m parts is equal to the number of partitions of $n + m$ with exactly m parts, i.e., prove that

$$\sum_{k=1}^{m} p_k(n) = p_m(n+m)$$

(a) by induction on m.

(b) by means of Ferrers diagrams.

14 Prove that

(a) $p_m(n) = p_m(n-m) + p_{m-1}(n-m) + \cdots + p_1(n-m), \quad m < n.$

(b) $p(n) = p_n(2n)$.

(c) $p(n) = p_{n+m}(2n+m), \quad m \geq 0.$

(d) For all $n \geq 8$, the last five (nonzero) numbers in row n of the partition triangle are 5, 3, 2, 1, 1.

(e) What is the generalization of Exercises 2(c) and 14(d)?

15 Suppose $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ and $\beta = [\beta_1, \beta_2, \ldots, \beta_k]$ are two partitions of n. Then α *majorizes* β if $m \leq k$ and

$$\sum_{i=1}^{r} \alpha_i \geq \sum_{i=1}^{r} \beta_i, \qquad 1 \leq r \leq m.$$

(a) Show that [6, 4] majorizes [4, 3, 2, 1].

(b) Show that [4, 3, 2, 1] majorizes $[3^2, 2^2]$.

(c) If α majorizes β and β majorizes γ, prove that α majorizes γ.

(d) Prove that α majorizes β if and only if β^* majorizes α^*.

16 Confirm that the coefficients 1, 6, 15, 20, 30, 60, and 90 in Example 1.8.17 are all correct.

17 Prove that the number of self-conjugate partitions of n is equal to the number of partitions of n that have distinct parts each of which is odd.

18 The great Indian mathematician Srinivasa Ramanujan (1887–1920) proved a number of theorems about partition numbers. Among them is the fact that $p(5n+4)$ is always a multiple of 5. Confirm this fact for $n = 0$, 1, and 2.

19 We saw in Section 1.6 that the equation $a+b+c+d+e=10$ has a total of $C(9,4)=126$ different positive integer solutions. Of these, how many satisfy $a \geq b \geq c \geq d \geq e$?

20 Denote by $t(n)$ the number of partitions of n each of whose parts is a power of 2 (including $2^0=1$).

(a) Compute $t(n), \quad 1 \leq n \leq 6$.

(b) Prove that $t(2n+1)=t(2n), \quad n \geq 1$.

(c) Prove that $t(2n)=t(n)+t(2n-2), \quad n \geq 2$.

(d) Prove that $t(n)$ is even, $n \geq 2$.

21 When $p(a,b,c,d)=(a+b+c+d)^{10}$ is expressed as a linear combination of minimal symmetric polynomials, compute the coefficient of

(a) $M_{[8,1^2]}(a,b,c,d)$. **(b)** $M_{[10]}(a,b,c,d)$.

(c) $M_{[3^2,2^2]}(a,b,c,d)$. **(d)** $M_{[3^2,2,1^2]}(a,b,c,d)$.

22 Compute the coefficient of

(a) $M_{[2,1^3]}(x_1,x_2,x_3,x_4,x_5,x_6)$ in $(x_1+x_2+x_3+x_4+x_5+x_6)^5$.

(b) $M_{[2,1^3]}(x_1,x_2,x_3,x_4,x_5)$ in $(x_1+x_2+x_3+x_4+x_5)^5$.

23 Express $p(x,y,z)$ as a linear combination of minimal symmetric polynomials, where

(a) $p(x,y,z)=5x^2+5y^2+5z^2-xy-xz-yz$.

(b) $p(x,y,z)=2x(1+2yz)-3x^2+2y-3y^2+2z-3z^2$.

24 Write out, in full,

(a) $M_{[5]}(w,x,y,z)$. **(b)** $M_{[4,1]}(w,x,y,z)$.

(c) $M_{[1^3]}(w,x,y,z)$. **(d)** $M_{[8,1]}(x,y,z)$.

(e) $M_{[3,2,1]}(x,y,z)$. **(f)** $M_{[3,1^2]}(x,y,z)$.

25 Theorem 1.8.15 can be use to custom design symmetric polynomials. The *homogeneous symmetric function* of degree n is defined by $H_0(x_1,x_2,\ldots,x_k)=1$ and

$$H_n(x_1,x_2,\ldots,x_k)=\sum_{\pi \vdash n} M_\pi(x_1,x_2,\ldots,x_k), \qquad n \geq 1,$$

where, recall, $M_\pi(x_1,x_2,\ldots,x_k)=0$ whenever $\ell(\pi)>k$. Explicitly write out all the terms in

(a) $H_2(x,y)$. **(b)** $H_3(x,y)$.

(c) $H_2(a,b,c)$. **(d)** $H_3(a,b,c)$.

26 Let $H_n(x_1,x_2,\ldots,x_k)$ be the homogeneous symmetric function defined in Exercise 25.

(a) Compare and contrast $H_n(x_1, x_2, \ldots, x_k)$ with $(x_1 + x_2 + \cdots + x_k)^n$. (*Hint:* See Equation (1.30).)

(b) Show that $H_n(x_1, x_2, \ldots, x_k)$ is the sum of $p_1(n) + p_2(n) + \cdots + p_k(n)$ different minimal symmetric polynomials.

(c) Prove that $H_n(x_1, x_2, \ldots, x_k)$ is the sum of $C(n + k - 1, n)$ different terms. (*Hint:* Theorem 1.7.5.)

(d) Prove that $H_n(x_1, x_2, \ldots, x_k) = H_n(x_1, x_2, \ldots, x_{k-1}) + x_k H_{n-1}(x_1, x_2, \ldots, x_k)$.

(e) Prove that $H_s(x_1, x_2, \ldots, x_n) - H_s(x_2, \ldots, x_n, x_{n+1}) = (x_1 - x_{n+1}) H_{s-1}(x_1, x_2, \ldots, x_{n+1})$.

27 Suppose m is a nonnegative integer. A *lattice path* of length m in the cartesian plane begins at the origin and consists of m unit "steps" each of which is either up or to the right. If s of the steps are up and $r = m - s$ of them are to the right, the path terminates at the point (r, s). "Directions" for the lattice path illustrated in Fig. 1.8.5 might go something like this: Beginning from (0, 0) (the lower left-hand corner), take two steps up, two to the right, one up, three right, one up, one right, and one up. If this grid were a street map and one were in the business of delivering packages, lattice paths would probably the called "routes", and these directions might be given in shorthand as UURRURRRURU. Suppose r and s are fixed but arbitrary nonnegative integers, with $r + s > 0$.

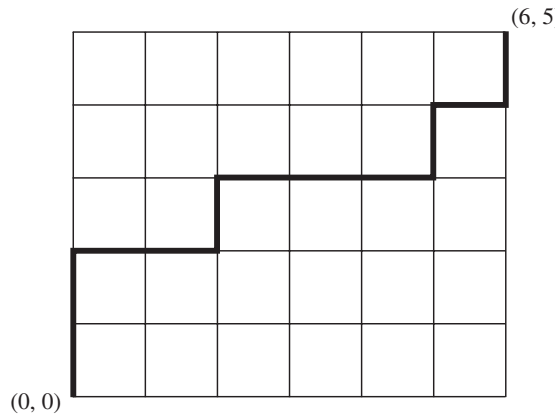


Figure 1.8.5

(a) Compute the number of different lattice paths from $(0, 0)$ to (r, s).

(b) The lattice path in Figure 1.8.5 "partitions" the 5×6 grid into two pieces. In this case, the piece above the path might easily be mistaken for the Ferrers diagram of partition $\pi = [6, 5, 2]$. Use this observation to compute the number of partitions that have at most s parts each of which is at most r.

(c) As an alternative to the alphabet $\{R, U\}$, one could just as well encode lattice paths using, say, the horizontal displacement of each step. In this scheme, each vertical step would correspond to a 0 and each horizontal step to a 1. For example, the lattice path in Fig. 1.8.5 would be encoded as the binary word 00110111010, a word of length 11 and "weight" 6. Compute the number of different binary words of length $r + s$ and weight r.

(d) Consider a binary word $w = b_1 b_2 \dots b_m$ of length m consisting of the letters (bits) $b_1, b_2, \dots, b_m$. The *inversion number* $\text{inv}(b_i) = 0$ if $b_i = 1$; if $b_i = 0$, it is the number of 1's to the left of b_i. If, e.g., $u = 00110111010$ (corresponding to Fig 1.8.5), the inversion numbers of its bits are 0, 0, 0, 0, 2, 0, 0, 0, 5, 0, and 6, respectively. In this case, the nonzero inversion numbers of u are precisely the parts of the corresponding partition π from part (b). Show that, in general, the nonzero inversion numbers of the bits of w are the parts of the partition to which w corresponds.

28 Galileo Galilei (1564–1642) once wondered about the frequency of throwing totals of 9 and 10 with three dice.

(a) Show that 9 and 10 have the same number of 3-part partitions each of whose parts is at most 6.

(b) Explain why it does not follow that 9 and 10 occur with equal frequency when three dice are rolled (repeatedly).

29 Suppose π is an m-part partition of n. Show that the number of different compositions of n that can be obtained by rearranging the parts of π is multinomial coefficient $\binom{n}{\pi}$.

1.9. ELEMENTARY SYMMETRIC FUNCTIONS

What immortal hand or eye could frame thy fearful symmetry?

— William Blake (*Songs of Experience*)

Let's begin by exploring the relationship between the coefficients of a monic polynomial

$$p(x) = x^n + c_1 x^{n-1} + c_2 x^{n-2} + \cdots + c_n \tag{1.31a}$$

and its roots $a_1, a_2, \dots, a_n$. Writing $p(x)$ in the form

$$p(x) = (x - a_1)(x - a_2) \cdots (x - a_n) \tag{1.31b}$$

suggests mimicking the alternative view of distributivity used to prove the binomial theorem, i.e., select one of x or $-a_1$ from the first set of parentheses, one of x or $-a_2$ from the second set, and so on. Finally, choose one of x or $-a_n$ from the nth set. String these selections together, in order, so as to create an n-letter "word", something like

$$(-a_1)xxx(-a_5)x\ldots xx.$$

If the total number of x's in this word is $n-r$, then the remaining "letters" are of the form $(-a_i)$ for r different values of i.

The sum of all such words is an inventory of the 2^n ways to make the sequence of decisions. Replacing each word with a monomial of the form

$$(-1)^r(a_1a_5\cdots)x^{n-r}$$

and combining terms of the same degree (in x) should yield Equation (1.31a). So, the coefficient of x^{n-r} in Equation (1.31a) must be the sum of all possible terms of the form

$$(-1)^r a_{i_1}a_{i_2}\cdots a_{i_r},$$

where $1 \le i_1 < i_2 < \cdots < i_r \le n$. In other words, c_r is $(-1)^r$ times the sum of the products of the roots taken r at a time. Let's give that sum a name.

1.9.1 Definition. The rth *elementary symmetric function*

$$E_r(x_1, x_2, \ldots, x_n)$$

is the sum of all possible products of r elements chosen from $\{x_1, x_2, \ldots, x_n\}$ without replacement where order doesn't matter.

Evidently, $E_r(x_1, x_2, \ldots, x_n)$ is the sum of all $C(n, r)$ "square-free" monomials of (total) degree r in the variables $x_1, x_2, \ldots, x_n$. Our conclusions about the relationship between roots and coefficients can now be stated as follows.

1.9.2 Theorem. *Let $a_1, a_2, \ldots, a_n$ be the roots of a monic polynomial $p(x) = x^n + c_1x^{n-1} + c_2x^{n-2} + \cdots + c_n$. Then*

$$c_r = (-1)^r E_r(a_1, a_2, \ldots, a_n), \qquad 1 \le r \le n. \tag{1.32}$$

1.9.3 Example. Suppose $f(x) = x^4 - x^2 + 2x + 2$. Then, counting multiplicities, $f(x)$ has four (complex) roots; call them a_1, a_2, a_3, and a_4. Setting $E_r = E_r(a_1, a_2, a_3, a_4)$ and comparing the actual coefficients of $f(x)$ with the generic

formula $f(x) = x^4 - E_1x^3 + E_2x^2 - E_3x + E_4$, we find that

$$\begin{aligned} 0 &= E_1(a_1, a_2, a_3, a_4) = a_1 + a_2 + a_3 + a_4, \\ -1 &= E_2(a_1, a_2, a_3, a_4) = a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 + a_2a_4 + a_3a_4, \\ -2 &= E_3(a_1, a_2, a_3, a_4) = a_1a_2a_3 + a_1a_2a_4 + a_1a_3a_4 + a_2a_3a_4, \\ 2 &= E_4(a_1, a_2, a_3, a_4) = a_1a_2a_3a_4. \end{aligned}$$

So, just from its coefficients, we can tell, e.g., that the sum of the roots of $f(x)$ is 0 and that their product is 2. □

1.9.4 Example. Suppose $a_i = 1$, $1 \le i \le n$, so that

$$\begin{aligned} p(x) &= (x-1)^n \\ &= C(n,0)x^n - C(n,1)x^{n-1} + C(n,2)x^{n-2} - \cdots + (-1)^n C(n,n). \end{aligned}$$

In this case, $E_r(1, 1, \ldots, 1) = C(n, r)$, $1 \le r \le n$, which makes perfect sense. After all, $E_r(a_1, a_2, \ldots, a_n)$ is the sum of all $C(n, r)$ products of the a_i's taken r at a time. If $a_i = 1$ for all i, then every one of these products is 1, and their sum is $E_r(1, 1, \ldots, 1) = C(n, r)$. □

Consistent with the fact that the leading coefficient of a monic polynomial is 1, we *define* $E_0(x_1, x_2, \ldots, x_n) = 1$.

1.9.5 Example. If $a_i = i$, $1 \le i \le 4$, then

$$\begin{aligned} &E_0(1,2,3,4) = 1, \\ &E_1(1,2,3,4) = 1 + 2 + 3 + 4 = 10, \\ &E_2(1,2,3,4) = 1\times 2 + 1\times 3 + 1\times 4 + 2\times 3 + 2\times 4 + 3\times 4 = 35, \\ &E_3(1,2,3,4) = 1\times 2\times 3 + 1\times 2\times 4 + 1\times 3\times 4 + 2\times 3\times 4 = 50, \\ &E_4(1,2,3,4) = 1\times 2\times 3\times 4 = 24. \end{aligned}$$

If $p(x) = (x-1)(x-2)(x-3)(x-4)$, then, with the abbreviation $E_r = E_r(1,2,3,4)$, $0 \le r \le 4$, Theorem 1.9.2 yields

$$\begin{aligned} p(x) &= E_0x^4 - E_1x^3 + E_2x^2 - E_3x + E_4 \\ &= x^4 - 10x^3 + 35x^2 - 50x + 24. \end{aligned}$$

Let's confirm this directly:

$$\begin{aligned} p(x) &= (x-1)(x-2)(x-3)(x-4) \\ &= (x^2 - 3x + 2)(x^2 - 7x + 12) \\ &= x^4 - (7+3)x^3 + (12 + 21 + 2)x^2 - (36 + 14)x + 24. \end{aligned}$$ □

Apart from their intrinsic significance, elementary symmetric functions have important (and, in some cases, unexpected) connections with other combinatorial

objects. Recall, e.g., that the number of ways to choose $n+1$ items from an m-element set without replacement where order matters is

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

1.9.6 Definition. The *falling factorial function* is defined by $x^{(0)} = 1$ and

$$x^{(n+1)} = x(x-1)(x-2)\cdots(x-n), \qquad n \geq 0.$$

Since $x^{(n+1)}$ is a polynomial of degree $n+1$, whose roots are $0, 1, \ldots, n$, and because $E_r(0, 1, \ldots, n) = E_r(1, 2, \ldots, n)$, $0 \leq r \leq n$, it follows that

$$\begin{aligned} x^{(n+1)} = {} & x^{n+1} - E_1(1, 2, \ldots, n)x^n + E_2(1, 2, \ldots, n)x^{n-1} - \cdots \\ & + (-1)^n E_n(1, 2, \ldots, n)x. \end{aligned}$$

In particular,

$$\begin{aligned} P(m, n+1) = {} & m\big[m^n - E_1(1, 2, \ldots, n)m^{n-1} + E_2(1, 2, \ldots, n)m^{n-2} - \cdots \\ & + (-1)^n E_n(1, 2, \ldots, n)\big]. \end{aligned}$$

Let's take a brief excursion* and investigate the numbers $E_t(1, 2, \ldots, n)$.

1.9.7 Definition. The *elementary number*

$$e(n, t) = \begin{cases} 0, & t < 0 \quad \text{or} \quad t > n, \\ E_t(1, 2, \ldots, n), & 0 \leq t \leq n. \end{cases}$$

Apart from Example 1.9.5, where we computed

$$\begin{aligned} (x-1)(x-2)(x-3)(x-4) &= x^4 - 10x^3 + 35x^2 - 50x + 24 \\ &= x^4 - e(4, 1)x^3 + e(4, 2)x^2 - e(4, 3)x + e(4, 4), \end{aligned}$$

we know that

$$\begin{aligned} e(n, 0) &= E_0(1, 2, \ldots, n) \\ &= 1, \\ e(n, 1) &= E_1(1, 2, \ldots, n) \\ &= 1 + 2 + \cdots + n \\ &= \frac{1}{2}n(n+1), \\ e(n, n) &= E_n(1, 2, \ldots, n) \\ &= 1 \times 2 \times \cdots \times n \\ &= n!. \end{aligned}$$

*There is a serious side to this excursion. In Chapter 2, we will discover that $s(n, r) = E_{n-r}(1, 2, \ldots, n-1)$ is a *Stirling number of the first kind*.

n \ t	0	1	2	3	4	5	6	7
1	1	1						
2	1	3	2					
3	1	6	$e(3,2)$	6				
4	1	10	35	50	24			
5	1	15	$e(5,2)$	$e(5,3)$	$e(5,4)$	120		
6	1	21	$e(6,2)$	$e(6,3)$	$e(6,4)$	$e(6,5)$	720	
7	1	28	$e(7,2)$	$e(7,3)$	$e(7,4)$	$e(7,5)$	$e(7,6)$	5040
				⋯				

Figure 1.9.1. Elementary triangle.

This gives us a start at filling in some entries of the *elementary triangle* exhibited in Fig. 1.9.1. What is (momentarily) missing is a recurrence for the elementary numbers analogous to Pascal's relation for binomial coefficients and/or to Theorem 1.8.7 for partition numbers.

1.9.8 Lemma. *If $n > t > 1$, then*

$$e(n,t) = e(n-1,t) + ne(n-1,t-1).$$

Proof: $E_t(1,2,\ldots,n) = e(n,t)$ is the sum of all $C(n,t)$ products of the numbers $1,2,\ldots,n$ taken t at a time. Some of these products involve n, and some do not. The sum of the products that do not involve n is $E_t(1,2,\ldots,n-1) = e(n-1,t)$. When n is factored out of the remaining terms, the other factor is $E_{t-1}(1,2,\ldots,n-1) = e(n-1,t-1)$. ■

From Fig. 1.9.1 and Lemma 1.9.8 we see, e.g., that

$$\begin{aligned} e(3,2) &= e(2,2) + 3e(2,1) \\ &= 2 + 3 \times 3 \\ &= 11. \end{aligned}$$

Similarly,

$$\begin{aligned} e(5,2) &= e(4,2) + 5e(4,1) \\ &= 35 + 5 \times 10 \\ &= 85, \end{aligned}$$

and

$$\begin{aligned} e(5,3) &= e(4,3) + 5 \times e(4,2) \\ &= 50 + 5 \times 35 \\ &= 225. \end{aligned}$$

Continuing in this way, a row at a time, one obtains Fig. 1.9.2.

n \ t	0	1	2	3	4	5	6	7
1	1	1						
2	1	3	2					
3	1	6	11	6				
4	1	10	35	50	24			
5	1	15	85	225	274	120		
6	1	21	175	735	1624	1764	720	
7	1	28	322	1960	6769	13132	13068	5040
				⋯				

Figure 1.9.2. The elementary numbers $e(n, t)$.

As their name implies, elementary symmetric functions are symmetric. Because multiplication is commutative, the coefficients of

$$p(x) = (x-1)(x-2)(x-3)(x-4)$$

are identical to the coefficients of

$$p(x) = (x-3)(x-1)(x-4)(x-2);$$

the sum of the products of $x_1, x_2, \ldots, x_n$ taken t at a time is equal to the sum of the products of any rearrangement of the x's, taken t at a time. In fact, elementary symmetric functions are minimal symmetric polynomials!

1.9.9 Theorem. *The tth elementary symmetric function is identical to the minimal symmetric polynomial corresponding to the partition* $[1^t]$*, i.e.,*

$$M_{[1^t]}(x_1, x_2, \ldots, x_n) = E_t(x_1, x_2, \ldots, x_n).$$

Proof. If $(r_1, r_2, \ldots, r_n)$ is some rearrangement of the sequene $(1, 1, \ldots, 1, 0, 0, \ldots, 0)$ consisting of t 1's followed by $n - t$ 0's, then

$$x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n} = x_{i_1} x_{i_2} \cdots x_{i_t},$$

where $1 \le i_1 < i_2 < \cdots < i_t \le n$, $r_{i_1} = r_{i_2} = \cdots = r_{i_t} = 1$, and the rest of the r's are zero. Adding the monomials corresponding to all possible rearrangements of $(1, 1, \ldots, 1, 0, 0, \ldots, 0)$ yields

$$M_{[1^t]}(x_1, x_2, \ldots, x_n) = \sum x_{i_1} x_{i_2} \cdots x_{i_t}, \tag{1.33}$$

where the sum is over $1 \le i_1 < i_2 < \cdots < i_t \le n$. In other words, the right-hand side of Equation (1.33) is the sum of all $C(n,t)$ products of the x's taken t and a time, which is the definition of $E_t(x_1, x_2, \ldots, x_n)$. ■

Conjugate to $[1^t]$ is the partition $[t]$.

1.9.10 Definition. The minimal symmetric polynomial corresponding to $[t]$ is the tth *power sum*, abbreviated

$$\begin{aligned} M_t(x_1, x_2, \ldots, x_n) &= M_{[t]}(x_1, x_2, \ldots, x_n) \\ &= x_1^t + x_2^t + \cdots + x_n^t. \end{aligned}$$

If $t = 1$, then

$$\begin{aligned} M_1(x_1, x_2, \ldots, x_n) &= x_1 + x_2 + \cdots + x_n \\ &= E_1(x_1, x_2, \ldots, x_n). \end{aligned} \tag{1.34}$$

Our interest in power sums goes back to Section 1.5, where it was discovered, e.g., that

$$\begin{aligned} M_1(1, 2, \ldots, n) &= 1 + 2 + \cdots + n \\ &= \tfrac{1}{2} n(n+1), \\ M_2(1, 2, \ldots, n) &= 1^2 + 2^2 + \cdots + n^2 \\ &= \tfrac{1}{6} n(n+1)(2n+1), \end{aligned} \tag{1.35}$$

$$\begin{aligned} M_3(1, 2, \ldots, n) &= 1^3 + 2^3 + \cdots + n^3 \\ &= \tfrac{1}{4} n^2 (n+1)^2, \end{aligned} \tag{1.36}$$

and so on.

Recall (Theorem 1.8.15) that a polynomial in n variables is symmetric if and only if it is a linear combination of minimal symmetric polynomials. In this sense, the minimal symmetric polynomials are building blocks from which all symmetric polynomials can be constructed. The power sums are also building blocks, but in a different sense. The following result is proved in Appendix A1.

1.9.11 Theorem.* *Any polynomial symmetric in the variables $x_1, x_2, \ldots, x_n$ is a polynomial in the power sums $M_t = M_t(x_1, x_2, \ldots, x_n)$, $1 \le t \le n$.*

*To be encountered in Section 3.6, the symmetric "pattern inventory" is a polynomial in the power sums. A description of that polynomial is the substance of Pólya's theorem.

1.9.12 Example. We do not need Theorem 1.9.11 to tell us that $p(x,y,z) = (x+y+z)^3$ as a polynomial in the power sums. By definition, $p(x,y,z) = M_1(x,y,z)^3$. What about something more interesting, like $M_{[2,1]}(x,y,z) = x^2y + x^2z + xy^2 + xz^2 + y^2z + yz^2$? Observe that the product

$$\begin{aligned} M_2(x,y,z)M_1(x,y,z) &= (x^2+y^2+z^2)(x+y+z) \\ &= x^3+y^3+z^3+x^2y+x^2z+xy^2+xz^2+y^2z+yz^2 \\ &= M_3(x,y,z) + M_{[2,1]}(x,y,z). \end{aligned}$$

So, $M_{[2,1]}(x,y,z) = M_2(x,y,z)M_1(x,y,z) - M_3(x,y,z)$. Similarly,

$$\begin{aligned} M_2(x,y,z)^2 &= (x^2+y^2+z^2)^2 \\ &= x^4+y^4+z^4+2x^2y^2+2x^2z^2+2y^2z^2 \\ &= M_4(x,y,z) + 2M_{[2,2]}(x,y,z), \end{aligned}$$

so that $M_{[2,2]}(x,y,z) = \frac{1}{2}[M_2(x,y,z)^2 - M_4(x,y,z)]$. □

1.9.13 Example. Let's see how to express elementary symmetric functions as polynomials in the power sums. Already having observed that $E_1(x,y,z) = M_1(x,y,z)$, consider $E_2(x,y,z) = xy + xz + yz$. Rearranging terms in

$$\begin{aligned} M_1(x,y,z)^2 &= (x+y+z)^2 \\ &= (x^2+y^2+z^2) + (2xy+2xz+2yz) \\ &= M_2(x,y,z) + 2E_2(x,y,z) \end{aligned}$$

yields

$$E_2(x,y,z) = \tfrac{1}{2}[M_1(x,y,z)^2 - M_2(x,y,z)]. \tag{1.37}$$

Similar computations starting from $M_1(x,y,z)^3 = (x+y+z)^3$ lead to the identity

$$E_3(x,y,z) = \tfrac{1}{6}[M_1(x,y,z)^3 - 3M_1(x,y,z)M_2(x,y,z) + 2M_3(x,y,z)]. \tag{1.38}$$

(Confirm it.) □

Surely, Equations (1.37) and (1.38) are examples of some more general relationship between power sums and elementary symmetric functions. To discover what that pattern is, let's return to the source. Suppose, e.g., that

$$\begin{aligned} p(x) &= (x-a_1)(x-a_2)\cdots(x-a_n) \\ &= x^n - E_1x^{n-1} + E_2x^{n-2} - \cdots + (-1)^nE_n, \end{aligned}$$

where $E_r = E_r(a_1, a_2, \ldots, a_n)$. Substituting $x = a_i$ in this equation yields

$$\begin{aligned} 0 &= p(a_i) \\ &= a_i^n - E_1a_i^{n-1} + E_2a_i^{n-2} - \cdots + (-1)^nE_n. \end{aligned}$$

Summing on i and setting $M_t = M_t(a_1, a_2, \ldots, a_n) = a_1^t + a_2^t + \cdots + a_n^t$, we obtain

$$0 = M_n - E_1M_{n-1} + E_2M_{n-2} - \cdots + (-1)^nnE_n,$$

the $t = n$ case of the following.

1.9.14 Newton's Identities. [*] For a fixed but arbitrary positive integer n, let $M_r = M_r(x_1, x_2, \ldots, x_n)$ and $E_r = E_r(x_1, x_2, \ldots, x_n)$. Then, for all $t \geq 1$,

$$M_t - M_{t-1}E_1 + M_{t-2}E_2 - \cdots + (-1)^{t-1}M_1E_{t-1} + (-1)^ttE_t = 0. \tag{1.39}$$

1.9.15 Example. The first four of Newton's identities are equivalent to

$$\begin{aligned} M_1 &= E_1, \\ M_2 - M_1E_1 &= -2E_2, \\ M_3 - M_2E_1 + M_1E_2 &= 3E_3, \\ M_4 - M_3E_1 + M_2E_2 - M_1E_3 &= -4E_4. \end{aligned}$$

The first identity, $M_1 = E_1$, is the same as Equation (1.34). Substituting M_1 for E_1 in the second identity yields

$$E_2 = \frac{1}{2}[M_1^2 - M_2], \tag{1.40}$$

extending to n variables and confirming Equation (1.37). Eliminating E_1 and E_2 from the third identity recaptures the following extension of Equation (1.38):

$$E_3 = \frac{1}{6}[M_1^3 - 3M_1M_2 + 2M_3]. \tag{1.41}$$

[*]Named for Isaac Newton (1642–1727).

Eliminating E_1, E_2, and E_3 from the fourth identity produces something new, namely,

$$E_4 = \frac{1}{24}[M_1^4 - 6M_1^2M_2 + 8M_1M_3 + 3M_2^2 - 6M_4]. \tag{1.42}$$

Evidently, Newton's identities can be used to express any elementary symmetric function as a polynomial in the power sums. □

Because $E_3(x_1, x_2) = 0$, the right-hand side of Equation (1.41) had better be zero when $n = 2$. Let's confirm that it is:

$$\begin{aligned} M_1^3 + 2M_3 &= (x_1 + x_2)^3 + 2(x_1^3 + x_2^3) \\ &= 3x_1^3 + 3x_1^2x_2 + 3x_1x_2^2 + 3x_2^3 \\ &= 3(x_1 + x_2)(x_1^2 + x_2^2) \\ &= 3M_1M_2. \end{aligned}$$

So, as predicted, $M_1^3 - 3M_1M_2 + 2M_3 = 0$. More generally, because $E_{n+r}(x_1, x_2, \ldots, x_n) = 0$, $r \geq 1$, Equation (1.39) has a simpler form when $t > n$, namely,

$$M_t - M_{t-1}E_1 + M_{t-2}E_2 - \cdots + (-1)^n M_{t-n}E_n = 0. \tag{1.43}$$

A proof of Newton's identities for all $t \geq 1$ can be found in Appendix A1.

1.9. EXERCISES

1 Without computing the roots of $f(x) = x^4 - x^2 + 2x + 2$, it was argued in Example 1.9.3 that their elementary symmetric functions are $E_1 = 0$, $E_2 = -1$, $E_3 = -2$, and $E_4 = 2$. Confirm this result by finding the four roots and then computing their elementary symmetric functions directly from the definition.

2 Show that $(a^2 + b^2) - (a + b)(a + b) + 2ab = 0$ (thus confirming the $n = t = 2$ case of Newton's identities).

3 Find the elementary symmetric functions of the roots of

(a) $x^4 - 5x^3 + 6x^2 - 2x + 1$. **(b)** $x^4 + 5x^3 + 6x^2 + 2x + 1$.

(c) $x^4 + 5x^3 - 6x^2 + 2x - 1$. **(d)** $2x^4 + 10x^3 - 12x^2 + 4x - 2$.

(e) $x^5 - x^3 + 3x^2 + 4x - 8$. **(f)** $x^5 + x^4 - 2x$.

4 Compute

(a) $E_t(1, 2, 3, 4, 5)$, $1 \leq t \leq 5$, directly from Definition 1.9.1 (*Hint:* Use row 5 of Fig. 1.9.2 to check your answers.)

(b) $E_5(1,2,3,4,5,6,7)$.

5 Find the missing coefficients in

(a) $x^{(5)} = x^5 - 10x^4 + 35x^3 - \_\_\_x^2 + \_\_\_x - \_\_\_$.

(b) $x^{(6)} = x^6 - \_\_\_x^5 + \_\_\_x^4 - 225x^3 + \_\_\_x^2 - \_\_\_x$.

6 Compute

(a) $E_3(1,2,3,4,5,6,7,8)$. **(b)** $E_4(1,2,3,4,5,6,7,8)$.

(c) $E_6(1,2,3,4,5,6,7,8)$. **(d)** $E_7(1,2,3,4,5,6,7,8)$.

7 Let $f(x) = b_0x^n + b_1x^{n-1} + \cdots + b_{n-1}x + b_n$ be a polynomial of degree n whose roots are $a_1, a_2, \ldots, a_n$. Prove that $E_t(a_1, a_2, \ldots, a_n) = (-1)^t b_t/b_0$.

8 Confirm that $6(abc + abd + acd + bcd) = M_1^3 - 3M_1M_2 + 2M_3$, where $M_t = a^t + b^t + c^t + d^t$, $1 \le t \le 3$.

9 Newton's identities were used in Equations (1.40)–(1.42) to express $E_t = E_t(x_1, x_2, \ldots, x_n)$ as a polynomial in the power sums $M_t = M_t(x_1, x_2, \ldots, x_n)$, $2 \le t \le 4$.

(a) Confirm by a direct computation that

$$a^2 + b^2 + c^2 + d^2 = E_1(a,b,c,d)^2 - 2E_2(a,b,c,d).$$

(b) Show that $M_2 = E_1^2 - 2E_2$ for arbitrary n.

(c) Express M_3 as a polynomial in elementary symmetric functions.

(d) Show that $M_4 = E_1^4 - 4E_1^2E_2 + 4E_1E_3 + 2E_2^2 - 4E_4$.

(e) Prove that any polynomial symmetric in the variables $x_1, x_2, \ldots, x_n$ is a polynomial in the elementary symmetric functions $E_t(x_1, x_2, \ldots, x_n)$, $1 \le t \le n$.*

10 Express the symmetric function $f(a,b,c,d)$ from Example 1.8.16 as a polynomial in power sums.

11 Express $x^3y + xy^3$ as a polynomial in

(a) $M_1(x,y)$ and $M_2(x,y)$. **(b)** $E_1(x,y)$ and $E_2(x,y)$.

12 Because equations like those in Exercises 9(b)–(d) are polynomial identities, any numbers can be substituted for the variables $x_1, x_2, \ldots, x_n$.

(a) Use this idea to show that $1^2 + 2^2 + \cdots + n^2 = e(n,1)^2 - 2e(n,2)$.

(b) Use Fig. 1.9.2 and the result of part (a) to evaluate $1^2 + 2^2 + 3^2 + 4^2 + 5^2$. (Confirm that your answer is consistent with Equation (1.35).)

(c) Find a formula for $1^3 + 2^3 + \cdots + n^3$ in terms of $e(n,t)$, $t \le n$. (*Hint:* Use your solution to Exercise 9(c).)

*This is the so-called *Fundamental Theorem of Symmetric Polynomials*.

(d) Use Fig. 1.9.2 and the result of part (c) to evaluate $1^3 + 2^3 + 3^3 + 4^3 + 5^3$. (Confirm that your answer is consistent with Equation (1.36).)

13 Let $E_t = E_t(a_1, a_2, \ldots, a_n)$, $0 \le t \le n$. Show that

(a) $(a_1 - 1)(a_2 - 1)\cdots(a_n - 1) = E_n - E_{n-1} + E_{n-2} - \cdots + (-1)^n E_0$.

(b) $(1 - a_1 x)(1 - a_2 x)\cdots(1 - a_n x) = E_0 - E_1 x + E_2 x^2 - \cdots + (-1)^n E_n x^n$.

14 If $n \ge t \ge 2$, prove that

$$E_t(a_1, a_2, \ldots, a_n) = E_t(a_1, a_2, \ldots, a_{n-1}) + a_n E_{t-1}(a_1, a_2, \ldots, a_{n-1}).$$

(*Hint:* See the proof of Lemma 1.9.8.)

15 Give the inductive proof that

$$\prod_{i=1}^{n}(x - a_i) = \sum_{t=0}^{n}(-1)^t E_t(a_1, a_2, \ldots, a_n)x^{n-t}.$$

16 If $f(x) = x^{(n+1)}$, show that $f'(0) = \pm n!$.

17 Show that

(a) $x^{(m+n)} = x^{(m)}(x - m)^{(n)}$.

(b) $(x + y)^{(n)} = \sum_{r=0}^{n} C(n, r)x^{(r)}y^{(n-r)}$.

18 Recall (Section 1.8, Exercise 15) that if $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ and $\beta = [\beta_1, \beta_2, \ldots, \beta_k]$ are two partitions of n, then α *majorizes* β if $m \le k$, and

$$\sum_{i=1}^{r}\alpha_i \ge \sum_{i=1}^{r}\beta_i, \qquad 1 \le r \le m.$$

(a) Show that majorization imposes a linear order on the $p_3(8) = 5$ partitions of 8 having three parts.

(b) Among the many properties of elementary symmetric functions is *Schur concavity*, meaning that $E_t(\alpha) \le E_t(\beta)$ whenever α majorizes β. Confirm this property for $2 \le t \le 3$ using the three-part partitions of 8.

(c) If you were to compute $E_3(\alpha)$ for each four-part partition α of 24, which partition would produce the maximum? (The minimum?)

19 Let $H_t = H_t(x_1, x_2, \ldots, x_n)$ be the homogeneous symmetric function of Section 1.8, Exercise 25. Then H_t is *Schur convex*, meaning that $H_t(\alpha) \ge H_t(\beta)$, whenever α majorizes β.

(a) Confirm this result for H_2 and the three-part partitions of 8.

(b) If you were to compute $H_4(\alpha)$ for each three-part partition α of 24, which partition would produce the maximum? (The minimum?)

20 Show that the general formula for E_t as a polynomial in the power sums M_t is $t!E_t = \det(L_t)$, where

$$L_t = \begin{pmatrix} M_1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ M_2 & M_1 & 2 & 0 & \cdots & 0 & 0 \\ M_3 & M_2 & M_1 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ M_{t-1} & M_{t-2} & M_{t-3} & M_{t-4} & \cdots & M_1 & t-1 \\ M_t & M_{t-1} & M_{t-2} & M_{t-3} & \cdots & M_2 & M_1 \end{pmatrix}.$$

(*Hint:* Use Cramer's rule on the following matrix version of Newton's identities:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ M_1 & -2 & 0 & 0 & \cdots \\ M_2 & -M_1 & 3 & 0 & \cdots \\ M_3 & -M_2 & M_1 & -4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ \vdots \end{pmatrix} = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \\ \vdots \end{pmatrix}.$$

21 Confirm that the result in Exercise 20, i.e., $t!E_t = \det(L_t)$, agrees with

(a) Equation (1.40) when $t = 2$.

(b) Equation (1.41) when $t = 3$.

(c) Equation (1.42) when $t = 4$.

22 Bertrand Russell* once wrote, "I used, when excited, to calm myself by reciting the three factors of $a^3 + b^3 + c^3 - 3abc$."

(a) Express $a^3 + b^3 + c^3 - 3abc$ as a product of *two* nontrivial polynomials that are symmetric in a, b, and c. (*Hint:* Example 1.9.15 and $M_1(a, b, c) = E_1(a, b, c)$.)

(b) Show that $(a + b + c)(a + \theta b + \theta^2 c)(a + \theta^2 b + \theta c) = a^3 + b^3 + c^3 - 3abc$, where $\theta = \frac{1}{2}(-1 + i\sqrt{3})$ is a *primitive cube root of unity.*

(c) Show that if $a^3 + b^3 + c^3 - 3abc$ is a product of three polynomials, each of which is symmetric in a, b, and c, then one (at least) of them is a constant polynomial.

23 Prove that

(a) $e(n, 2) = C(n + 1, 2)$.

(b) $e(n, 3) = \frac{1}{48}(n - 2)(n - 1)n^2(n + 1)^2$.

*In 1914, having completed *Principia Mathematica* with Alfred North Whitehead, Bertrand Russell (1872–1970), Third Earl Russell, abandoned mathematics in favor of philosophy, social activism, and writing. He was awarded the Nobel Prize for Literature in 1950.

24 Show that $\{x^{(n)} : 0 \le n \le m\} = \{1, x, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$ is a basis for the vector space of polynomials of degree at most m. (*Hint:* Show that any polynomial $f(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$ of degree at most m can be expressed (uniquely) as a linear combination of $1, x, x^{(2)}, x^{(3)}, \ldots, x^{(m)}$.)

25 Let A be a real, symmetric, $n \times n$ matrix with characteristic polynomial

$$\det(xI_n - A) = x^n - c_1 x^{n-1} + c_2 x^{n-2} - \cdots + (-1)^n c_n.$$

Show that

(a) $c_1 = \sum_{i=1}^{n} a_{ii} = tr(A)$, the *trace* of A.

(b) $c_2 = \frac{1}{2}[\text{tr}(A)^2 - \text{tr}(A^2)]$

(c) $c_3 = \frac{1}{6}[\text{tr}(A)^3 - 3\ \text{tr}(A)\ \text{tr}(A^2) + 2\ \text{tr}(A^3)]$

(d) $\text{tr}(A^t) - c_1\,\text{tr}(A^{t-1}) + c_2\,\text{tr}(A^{t-2}) - \cdots + (-1)^t t c_t = 0,\ t \ge 1.$

26 Recall that $[k, 1^m]$ is shorthand for the partition of $m + k$ consisting of a single k followed by m 1's.

(a) Show that $M_s(x_1, x_2, \ldots, x_n)E_t(x_1, x_2, \ldots, x_n) = M_{[s+1,1^{t-1}]}(x_1, x_2, \ldots, x_n) + M_{[s,1^t]}(x_1, x_2, \ldots, x_n)$, $s > 1$.

(b) Show that $M_1(x_1, x_2, \ldots, x_n)E_t(x_1, x_2, \ldots, x_n) = M_{[2,1^t]}(x_1, x_2, \ldots, x_n) + (t+1)E_{t+1}(x_1, x_2, \ldots, x_n)$.

(c) Base a proof of Newton's identities on parts (a) and (b).

*1.10. COMBINATORIAL ALGORITHMS

> In a few generations you can bread a racehorse. The recipe for making a man like Delacroix is less well known.
>
> — Jean Renoir

Algos is the Greek word for "pain"; *algor* is Latin for "to be cold"; and Al Gore is a former Vice President of the United States. Having no relation to any of these, *algorithm* derives from the ninth-century Arab mathematician Mohammed ben Musa al-Khowârizmi.* Translated into Latin in the twelfth century, his book *Algorithmi de numero Indorum* consists of step-by-step procedures, or recipes, for solving arithmetic problems.

As an illustration of the role of algorithms in mathematics, consider the following example: one version of the *well-ordering principle* is that any nonempty set of

*Mohammed, son of Moses, of Khowârizm. Al-Khowârizmi also wrote *Hisâb al-jabr wa'1 muqâbalah*; from which the word *algebra* is derived. It was largely through the influence of his books that the Hindu-Arabic numeration system reached medieval Europe.

positive intergers contains a least element. Given two positive integers a and b, well ordering implies the existence of a least element d of the set

$$\{sa + tb : s \text{ and } t \text{ are integers and } sa + tb > 0\}.$$

This least element has a name; it is the greatest common divisor (GCD) of a and b. Well ordering establishes the existence of d but furnishes little information about its value. For that we must look elsewhere.

Among the algorithms for computing GCDs is one attributed to Euclid, based on the fact that if r is the remainder when a is divided by b, then the GCD of a and b is equal to the GCD of b and r. A different algorithm is based on the unique prime factorizations of a and b. Either algorithm works just fine for small numbers, where the second approach may even have a conceptual advantage. For actual computations with large numbers, however, the Euclidean algorithm is much easier and much *much* faster.

Not until digital computers began to implement algorithms in calculations involving astronomically large numbers did the mathematical community, *as a whole*, pay much attention to these kinds of computational considerations. Courses in the analysis of algorithms are relatively new to the undergraduate curriculum.

This section is devoted to a naive introduction to a few of the ideas associated with combinatorial algorithms. Let's begin with the multinomial coefficient

$$M = \binom{n}{r_1, r_2, \ldots, r_k}$$

$$= \frac{n!}{r_1! r_2! \cdots r_k!},$$

where, e.g.,

$$n! = 1 \times 2 \times \cdots \times n.$$

Observe that $n!$ is not so much a number as an algorithm for computing a number. To compute $n!$, multiply 1 by 2, multiply their product by 3, multiply that product by 4, and so on, stopping only when the previous product has been multiplied by n.

The following is a subalgorithm, or *subroutine*, to compute the factorial F of an arbitrary integer X:

1. Input X.
2. $F = 1$ and $I = 0$.
3. $I = I + 1$.
4. $F = F \times I$.
5. If $I < X$, then go to step 3.
6. Return F.

These lines should be interpreted as a step-by-step recipe that, absent directions to the contrary (like "go to step 3"), is to be executed in numerical order. In step 6, the value *returned* is $F = X!$.

This subroutine is written in the form of a primitive computer program. To a hypothetical computer, symbols like X, F, and I are names for memory locations. Step 1 should be interpreted as an instruction to wait for a number to be entered, then to store the number in some ("random"[*]) memory location and, so as not to forget the location, flag it with the symbol X. In step 2, the numbers 1 and 0 are stored in memory locations labeled F and I, respectively. In step 3, the number in memory location I is replaced with the next larger integer.[†] In step 4, the number in memory location F is replaced with the product of the number found there, and the number currently residing in memory location I. If, in step 5, memory location I contains X, operation moves on to step 6, where the subroutine terminates by returning $F = X!$. Otherwise, the action loops back to step 3 for another iteration.

The *loop* in steps 3–5 can be expressed more compactly using the equivalent "For ... Next" construction foud in steps 3–5 of the following:

1.10.1 (Factorial Subroutine) Algorithm

1. Input X.
2. $F = 1$.
3. For $I = 1$ to X.
4. $F = F \times I$.
5. Next I.
6. Return F. □

The factorial subroutine affords the means to compute $n!$, $r_1!$, $r_2!$, and so on, from which the multinomial coefficient $M = \binom{n}{r_1, r_2, \ldots, r_k}$ can be obtained, either as the quotient of $n!$ and the product of the factorials of the r's or, upon dividing $n!$ by $r_1!$, dividing the quotient by $r_2!$, dividing that quotient by $r_3!$, and so on. While these two approaches may be arithmetically equivalent, they represent *different* algorithms.

1.10.2 (Multinomial Coefficient) Algorithm

1. Input $n, k, r_1, r_2, \ldots, r_k$.
2. $X = n$.
3. Call Algorithm 1.10.1.
4. $M = F$.
5. For $j = 1$ to k.
6. $X = r_j$

[*]Hence the name random-access memory, or RAM.

[†]Notations such as "$I \leftarrow I + 1$" or "$I := I + 1$" are sometimes used in place of "$I = I + 1$".

```
 7. Call Algorithm 1.10.1.
 8. M=M/F.
 9. Next j.
10. Return M.
```

□

Having let $X = n$ in step 2, the factorial subroutine is called upon in step 3 to return $F = n!$. Thus, in step 4, the number entered into memory location M is $n!$. On the first trip through the loop in steps 5–9, $j = 1$ and $X = r_1$. When the factorial subroutine is called in step 7, the number it returns is $F = r_1!$ so, in step 8, the number in memory location M is replaced by $n!/r_1!$. Assuming $j < k$ in step 9, action is directed back to step 5, and the value of j is increased by 1. The second time step 8 is encountered, the number currently being stored in memory location M, namely, $n!/r_1!$, is replaced with $(n!/r_1!)/r_2! = n!/(r_1!r_2!)$. And so on. Finally, the kth and last time step 8 is encountered, the number in memory location M is replaced with $\binom{n}{r_1,r_2,\ldots,r_k}$.

It might be valuable to pause here and give this algorithm a try, either by writing a computer program to implement it or by following the steps of Algorithm 1.10.2 yourself as if you were a (*virtual*) computer. Test some small problem, the answer to which you already know, e.g., $\binom{11}{4,4,2,1} = 34,650$ from the original MISSISSIPPI problem. After convincing yourself that the algorithm works properly, try it on $C(100, 2)$.

Whether your computer is virtual or real, using Algorithm 1.10.2 to compute $C(100, 2)$ may cause it to choke. If this happens, the problem most likely involves the magnitude of 100!. The size of this number can be estimated by means of an approximation known as *Stirling's formula*[*]:

$$n! \doteq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{1.44}$$

Using common logarithms, $100/e = 36.8 \doteq 10^{1.57}$, so $(100/e)^{100} \doteq 10^{157}$. Since $\sqrt{2\pi} \times 10 \doteq 25$, Equation (1.44) yields $100! \doteq 2.5 \times 10^{158}$. (Current estimates put the age of the universe at something less than 5×10^{26} nanoseconds.)

Without a calculator or computer, one would not be likely even to consider evaluating $C(100, 2)$ by first computing 100!, because something along the following lines is so much easier:

$$\begin{aligned} C(100, 2) &= \frac{98! \times 99 \times 100}{98! \times 1 \times 2} \\ &= 99 \times 50 \\ &= (100 - 1) \times 50 \\ &= 4950. \end{aligned}$$

[*]Stirling's formula should not be confused with *Stirling's identity*, soon to be encountered in Chapter 2.

The key to converting this easier approach into an algorithm is best illustrated with a slightly less trivial example, e.g., (see Theorem 1.5.1)

$$\binom{n}{r,s,t} = \frac{P(n,r)}{r!} \times \frac{P(n-r,s)}{s!} \times \frac{P(n-r-s,t)}{t!}. \tag{1.45}$$

Viewing $P(n,r)/r!$ as

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

$P(n-r,s)/s!$ as

$$\frac{n-r}{1} \times \frac{n-r-1}{2} \times \cdots \times \frac{n-r-s+1}{s},$$

and so on, suggests another subroutine:

```
1. M = 1.
2. For J = 1 to r.
3. M = M × N/J.
4. N = N - 1.
5. Next J.
```

Setting $N = n$ and $r = r_I$ and *nesting* this subroutine inside a "For $I = 1$ to k" loop yields another algorithm.

Can we do better? Almost surely. Because $n = r + s + t$, the last factor in Equation (1.45) is $P(t,t)/t! = t!/t! = 1$. Evidently, "For $I = 1$ to $k - 1$" suffices in the "outside loop". On the other hand, since $\binom{n}{r_1,r_2,\ldots,r_k} = \binom{n}{r_2,\ldots,r_k,r_1}$, the outside loop could just as well be "For $I = 2$ to k".

1.10.3 (Improved Multinomial Coefficient) Algorithm

```
1. Input n, k, r1, r2, ..., rk.
2. M = 1 and N = n.
3. For I = 2 to k.
4. For J = 1 to rI.
5. M = M × N/J.
6. N = N - 1.
7. Next J.
8. Next I.
9. Return M.
```

□

LUCK	LUKC	LCUK	LCKU	LKUC	LKCU
ULCK	ULKC	UCLK	UCKL	UKLC	UKCL
CLUK	CLKU	CULK	CUKL	CKLU	CKUL
KLUC	KLCU	KULC	KUCL	KCLU	KCUL

Figure 1.10.1. Rearrangements of LUCK.

It is clear from our experience so far that different algorithms can achieve the same outcome, *some better than others*! Algorithm 1.10.3 is superior to Algorithm 1.10.2 because it is more widely applicable. (Check to see that calculating $C(100, 2)$ is no trouble for Algorithm 1.10.3.) In general, however, it is not always clear which of two (or more) algorithms is best. It may not even be clear how to interpret "best"!

This book began with a discussion of the four-letter words that can be produced by rearranging the letters in LUCK. An initial (brute-force) approach resulted in a systematic list, reproduced in Fig. 1.10.1 for easy reference. In subsequent discussions, it was often useful to *imagine* constructing a list, with the implied understanding that list making is mildly distasteful. And, so it is, as long as the only reason to make a list is to count the words on it! Such peremptory judgments do not apply when the list serves other purposes. There are, in fact, many good reasons to make a list.

Suppose one had a reason for wanting a list of the $4! = 24$ rearrangements of LUCK, e.g., to use in constructing a master list of encryption keys upon which to base monthly corporate passwords for the next two years. In order to be most useful, such a list should be organized so that specific words are easy to locate. Figure 1.10.1 gives one possibility, based on the order in which the letters appear in LUCK. A more common approach is based on the order in which letters appear in the alphabet.

1.10.4 Definition. Let $X = x_1x_2 \ldots x_p$ and $Y = y_1y_2 \ldots y_q$ be words containing p and q letters, respectively. Then X comes before Y, in *dictionary order*,* if x_1 comes before y_1 in alphabetical order; or if there is a positive integer $r \leq p$ such that $x_i = y_i$, $1 \leq i < r$, and x_r precedes y_r in alphabetical order; or if $p < q$ and $x_i = y_i$, $1 \leq i \leq p$.

A list of words in dictionary order is often called an *alphabetized list*, and dictionary order is sometimes referred to as "alphabetical order." Whatever such lists are called, algorithms to generate them are surprisingly difficult to design. Our approach takes advantage of the numerical order that is already hard-wired into computers.

* Dictionary order is also known as *lexicographic order*, *lexicon* being another word for "dictionary".

1.10.5 Example. Consider "*words*" assembled from the *alphabet* $\{0, 1, 2, \ldots, 9\}$. Suppose *alphabetical order* for these ten "*letters*" is interpreted as numerical order. Would it surprise you to learn that, in this context, dictionary order does not coincide with the usual extension of numerical order? While 9 comes before 10 in numerical order, 9 comes after 10 in dictionary order! (Confirm that, upon restriction to number/words of the same length, the two orderings *do* coincide.) □

1.10.6 Example. In the spirit of Example 1.10.5, consider the $4! = 24$ four-letter words that can be assembled by rearranging the letters/digits in 3142. Among the challenges that stand between us and an algorithm to generate and list these words in dictionary order is familiarity! We do chores like this all the time without thinking about *how* we do them.

Let's start at the beginning, focusing on process: Since 1 comes first in alphabetical order, any word that begins with 1 will precede, in dictionary order, all words that begin with something else. Similarly, among the words whose first letter is 1, any whose second letter is 2 will precede all those whose second letter is not. Continuing in this way, it is easy to see that the list must begin with 1234, the unique rearrangement of 3142 in which the letters occur in increasing alphabetical order. Reversing the argument shows that the last word on the list is 4321, the unique word in which the letters decrease, in alphabetical order (when read from left to right).

Because only two rearrangements of 3142 have initial fragment 12, the word following 1234 on the list can only be 1243. Indeed, any two words with the same initial fragment have tailing fragments consisting of the same (complementary) letters. Moreover, all words with the same initial fragment must appear consecutively on the list, starting with the word in which the tailing letters are arranged in increasing order and ending with the word in which the tailing letters are in decreasing order.

After 1243 come the words with initial fragment 13. In the first of these, the tail is 24, and in the second it is 42. The observation that 42 is the reverse of 24 suggests a two-step procedure for finding the next word after 1342 on the list.

In the first step, 1342 is transformed into the intermediate word 1432 by switching the positions of 3 and 4. Observe that, while the switch changes the tail from 42 to 32, the new tail is (still) in decreasing order. In the second step, this intermediate word is transformed from last to first among the words with initial fragment 14 by reversing its tail. The result, 1423, is the next rearrangement of 3142 after 1342.

What comes after 1423? Well, 1432, of course! But, how does 1432 emerge from the two-step process outlined in the previous paragraph? Because 1423 is the only word on the list that begins with 142, it is the last word on the list with initial fragment 142. (This time, the tail is 3.) Switching 2 and 3 results in the intermediate word 1432 (whose tail is 2). Because a tail of length one reverses to itself, the output of the two-step process is 1432.

What comes after 1432? Because 432 is in decreasing order, 1432 is the last word on the list with initial fragment 1. Switching 1 with 2 produces the intermediate word 2431. Reversing the tail, 431, yields the next word on the list, namely, 2134.

Imagine yourself somewhere in the middle of the list, having just written the word $d_1d_2d_3d_4$. Using the two-step process to find the next word depends on being able to recognize the letter to be switched. The key to doing that is the tail. Assuming $d_1d_2d_3d_4 \neq 4321$, the only way it can be the last word on the list with initial fragment $d_1 \ldots d_j$ is if letters $d_{j+1}, \ldots, d_4$ are in decreasing order. For d_j to be the letter that gets switched, there must be some letter in the tail with which to switch it, i.e., some $d_k \in \{d_{j+1}, \ldots, d_4\}$ that comes after d_j in alphabetical (numerical) order. If $d_j, d_{j+1}, \ldots, d_4$ were in decreasing order, there could be no such d_k.

In the two-step process, the tail is the longest fragment (starting from the right-hand end of $d_1d_2d_3d_4$) whose letters decrease (when read from left to right). Put another way, the letter to be switched is d_j, where j is the largest value of i such that $d_i < d_{i+1}$. Once j has been identified, step 1 is accomplished by switching d_j with d_k, where d_k is the smallest letter in the tail that is larger than d_j, i.e.,

$$d_k = \min\{d_i : i > j \text{ and } d_i > d_j\}. \tag{1.46}$$

(Because $d_{j+1} > d_j$ and because d_{j+1} belongs to the tail, d_k always exists.)

When d_j and d_k are switched, a new tail is produced in which d_k (from the old tail) has been replaced by d_j. Because of the way j and d_k have been chosen, the letters in the new tail are (still) decreasing. Reversing the new tail in step 2 is equivalent to rearranging its letters into increasing order. □

The discussion in Example 1.10.6 leads to an algorithm for listing, in dictionary order, all rearrangements of 3142.

1.10.7 Algorithm

1. Set $d_i = i$, $1 \leq i \leq 4$.
2. Write $d_1\, d_2\, d_3\, d_4$.
3. If $d_i > d_{i+1}$, $1 \leq i \leq 3$, then stop.
4. Let j be the largest i such that $d_i < d_{i+1}$.
5. Let k be chosen to satisfy Equation (1.46).
6. Switch d_j and d_k.*
7. Reverse $d_{j+1}, \ldots, d_4$.
8. Go to step 2. □

It would not be a bad idea to pause and implement Algorithm 1.10.7 on a computer (real or virtual) and check to see that the output is something closely resembling Fig. 1.10.2.

What about the master list of encryption keys upon which to base monthly corporate passwords for the next two years? An algorithm to generate a list, in

*So that the new d_j is the old d_k, and vice versa.

1234	1243	1324	1342	1423	1432
2134	2143	2314	2341	2413	2431
3124	3142	3214	3241	3412	3421
4123	4132	4213	4231	4312	4321

Figure 1.10.2. The 24 rearrangements of 1234.

dictionary order, of all 24 rearrangements of LUCK, is only a step or two from Algorithm 1.10.7. The missing steps involve explaining to a computer that C, K, L, U is an alphabetical listing of the letters in LUCK.* This is most easily accomplished using ''string variables''.

Like a word, a text *string* is a sequence (ordered concatenation) of symbols. Like numbers, strings of text can be stored in memory locations and labeled with symbols. But, it is often necessary to choose labels that distinguish string memory locations from those used to store numbers. We will use a dollar sign to indicate a string variable. The notation $A\$(4) =$ ''FOOD'', e.g., indicates that the string FOOD should be stored in the fourth cell of an *array* of string variable memory locations labeled $A\$$.

1.10.8 Example. To convert Algorithm 1.10.7 to an algorithm for generating, in dictionary order, the rearrangements of LUCK, add step

0. $L\$(1) =$ "C", $L\$(2) =$ "K", $L\$(3) =$ "L", $L\$(4) =$ "U"

and modify step 2 so that it reads

2. Write $L\$(d_1)L\$(d_2)L\$(d_3)\ L\$(d_4)$. □

Why not pause, modify Algorithm 1.10.7 now, and confirm that its output resembles Fig. 1.10.3. (Compare with Fig. 1.10.1.)

1.10.9 Example. The conversion of Algorithm 1.10.7 in Example 1.10.8 was relatively easy because the letters L, U, C, and K are all different. How much harder would it be to design an algorithm to generate, in dictionary order, all $4!/2 = 12$ four-letter rearrangements of LOOK?

CKLU	CKUL	CLKU	CLUK	CUKL	CULK
KCLU	KCUL	KLCU	KLUC	KUCL	KULC
LCKU	LCUK	LKCU	LKUC	LUCK	LUKC
UCKL	UCLK	UKCL	UKLC	ULCK	ULKC

Figure 1.10.3. Rearrangements of LUCK in dictionary order.

*As the name *digital computer* suggests, these machines were conceived and designed to crunch numbers. Numerical order is programmed into their genes, so to speak. Tasks related to word processing, on the other hand, have to be ''learned'', or ''memorized'' (which is why word processing software takes up so much space on a hard drive).

Let's begin with an algorithm to produce, in dictionary order, all twelve rearrangements of 1233. This is surprisingly easy! It can be done by replacing step 1 in Algorithm 1.10.7 with

$$\texttt{1. Set } d_1 = 1\texttt{, } d_2 = 2\texttt{, } d_3 = 3\texttt{, and } d_4 = 3$$

and replacing "$<$" in step 4 with "$\leq$".

To generate an ordered list of the rearrangements of LOOK, it suffices to modify this modified algorithm in the same way that Algorithm 1.10.7 was modified to obtain Example 1.10.8, namely, by adding step

$$\texttt{0. } L\$(1) = \text{“K”}\texttt{, } L\$(2) = \text{“L”}\texttt{, } L\$(3) = \text{“O”}$$

and changing step 2 to

$$\texttt{2. Write } L\$\ (d_1)\ L\$\ (d_2)\ L\$\ (d_3)\ L\$\ (d_4)\texttt{.}$$

At this point, how hard can it be to write an algorithm for listing, in dictionary order, all 11-letter words that can be produced by rearranging the letters in MISSISSIPPI? □

It is one thing to generate and list, in dictionary order, all possible rearrangements of the letters in some arbitrary word. It is something else to rearrange some arbitrary list of words into dictionary order. The latter is a so-called *sorting problem*. The comparison of various sorting algorithms affords a natural introduction to some applications of combinatorics in the analysis of algorithms. Those interested in pursuing such a discussion are referred to Appendix A2.

1.10.10 Example. A systematic listing of the seven partitions of 5 might be expected to look like this:

$$[5], [4, 1], [3, 2], [3, 1, 1], [2, 2, 1], [2, 1, 1, 1], [1, 1, 1, 1, 1].$$

In *reverse* dictionary order, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_\ell] \vdash n$ comes before $\beta = [\beta_1, \beta_2, \ldots, \beta_s] \vdash n$ if (and only if) $\alpha_1 > \beta_1$ or there is an integer $t < \ell$ such that $\alpha_i = \beta_i$, $1 \leq i \leq t$, and $\alpha_{t+1} > \beta_{t+1}$. Let's see if we can devise an algorithm to generate and list, in reverse dictionary order, all $p(n)$ partitions of n.

Because the list begins with $[n]$, all that's required is a step-by-step procedure to find the next partition, in reverse dictionary order, after a fixed but arbitrary $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_\ell] \neq [1^n]$ (the last partition on the list). There are two cases.

Case 1: If $\alpha_\ell = 1$, then $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_k, 1, \ldots, 1]$, where 1 occurs with multiplicity m, $\alpha_k > 1$, and $\ell = k + m$. If μ is the next partition after α, then μ is the first partition, in reverse dictionary order, that satisfies the conditions $\mu_i = \alpha_i$, $1 \leq i < k$,

and $\mu_k = \alpha_k - 1$. To find μ, let $S = \alpha_k + m$, the sum of the parts of α coming after α_{k-1}. If q is the quotient and r the remainder, when S is divided by $d = \alpha_k - 1$, then

$$\mu = [\alpha_1, \alpha_2, \ldots, \alpha_{k-1}, \alpha_k - 1, \ldots, \alpha_k - 1, r],$$

where $\alpha_k - 1$ occurs with multiplicity q and it is understood that r does not appear if it is zero.

Case 2: If $\alpha_\ell > 1$, the next partition after α is

$$\mu = [\alpha_1, \alpha_2, \ldots, \alpha_{\ell-1}, \alpha_\ell - 1, 1]. \qquad \square$$

Let's design an algorithm to implement the ideas of Example 1.10.10. Suppose

$$\alpha = [n^{m(n)}, \ldots, 2^{m(2)}, 1^{m(1)}],$$

where $i^{m(i)}$ is understood not to appear when $m(i) = 0$. If $m(1) = n$, then $\alpha = [1^n]$ and the list is complete. Otherwise, let j be the smallest integer larger than 1 such that $m(j) > 0$. The steps used in Example 1.10.10 to produce μ, the next partition after α, are these. Replace $m(j)$ with $m(j) - 1$. In case 1 (the case in which $m(1) > 0$), let q and r be the quotient and remainder when $S = j + m(1)$ is divided by $d = j - 1$. Set $m(1) = 0$; then set $m(j-1) = q$ and, if $r > 0$, set $m(r) = 1$. In case 2, if $j = 2$, set $m(1) = 2$; otherwise, set $m(j-1) = 1$ and $m(1) = 1$. A formal algorithm might look like this:

1.10.11 (Partition Generating) Algorithm

1. Input n.
2. Set $m(i) = 0$, $1 \le i < n$, and $m(n) = 1$.
3. Write $[n^{m(n)}, \ldots, 2^{m(2)}, 1^{m(1)}]$.
4. If $m(1) = n$, then stop.
5. $S = m(1)$.
6. $m(1) = 0$.
7. $j = 1$.
8. $j = j + 1$.
9. If $m(j) = 0$, then go to step 8.
10. $D = j - 1$.
11. $m(j) = m(j) - 1$.
12. If $S = 0$, then go to step 19.
13. $S = S + j$.
14. $Q = \lfloor S/D \rfloor$.
15. $R = S - D \times Q$.
16. $m(D) = Q$.
17. If $R > 0$, then $m(R) = 1$.

```
18. Go to step 3.
19. If j = 2, then go to step 23.
20. m(D) = 1.
21. m(1) = 1.
22. Go to step 3.
23. m(1) = 2.
24. Go to step 3.
```
□

Note that case 1 is addressed in steps 13–18 of Algorithm 1.10.11, while case 2 is handled in steps 19–24.

Having endured the development of Algorithm 1.10.11, why not convert it to a computer program and have the satisfaction of seeing the partitions of n appear on a computer screen?

1.10. EXERCISES

1 Write an algorithm to list the integers 1–100 in numerical order.

2 Write an algorithm to input two numbers and output

(a) their product.

(b) their sum.

(c) their difference.

3 Assuming that $r_1, r_2, \ldots, r_k$ vary in size, which of them should be chosen to play the role of r_1 in Algorithm 1.10.3?

4 Without actually running any programs, describe the output that would be produced if step 0 in Example 1.10.8 were replaced with

```
0. L$(1) = "K", L$(2) = "L", L$(3) = "O", L$(4) = "O".
```

5 Write an algorithm to generate and list, in dictionary order,

(a) all $5! = 120$ rearrangements of LUCKY.

(b) all $4!/2 = 12$ rearrangements of COOL.

6 Write an algorithm to compute and output the first ten rows (as n goes from 0 to 9) of Pascal's triangle. Base your algorithm on

(a) the algebraic formula $C(n, r) = n!/[r!(n - r)!]$.

(b) Pascal's relation.

7 Write an algorithm to generate and list, in dictionary order, all rearrangements of

(a) BANANA. **(b)** MISSISSIPPI. **(c)** MATHEMATICS.

8 Write an algorithm to generate and output the first ten rows of the partition triangle (i.e., the array whose (n, m)-entry is $p_m(n)$, the number of m-part partitions of n).

9 Write an algorithm to input n and output $p(n)$, the number of partitions of n. Base your algorithm on

(a) your solution to Exercise 8.

(b) Algorithm 1.10.11.

10 Write an algorithm to input a_0–a_4 and b_0–b_3 and to output the coefficient of x^k, $7 \geq k \geq 0$, in the product

$$(a_0x^4 + a_1x^3 + \cdots + a_4)(b_0x^3 + b_1x^2 + \cdots + b_3).$$

11 Write an algorithm to input x_1–x_6 and to output

(a) the third elementary symmetric function, $E_3(x_1, x_2, \ldots, x_6)$.

(b) all C(6,3) three-element subsets of $\{1, 2, 3, 4, 5, 6\}$.

(c) all $C(6, 3)$ three-element subsets of $\{x_1, x_2, \ldots, x_6\}$.

12 Write an algorithm to input x_1–x_6 and to output

(a) $E_2(x_1, x_2, \ldots, x_6)$.

(b) all $C(6, 2)$ two-element subsets of $\{x_1, x_2, \ldots, x_6\}$.

(c) the complements of the subsets in part (b).

(d) $E_4(x_1, x_2, \ldots, x_6)$.

13 Write an algorithm to input six *positive* numbers x_1–x_6 and to output $E_5(x_1, x_2, \ldots, x_6)$.

14 Write an algorithm to input the parts of a partition and output the parts of its conjugate.

15 Assuming 0 comes before 1 in alphabetical order, write an algorithm to generate and output, in dictionary order,

(a) all binary words of length 4 (i.e., all four-letter words that can be assembled using the alphabet $\{0, 1\}$).

(b) all binary words of length 8 and weight 4, where the weight of a binary word is the number of 1's among its bits.

16 Write an algorithm to input n and output, in dictionary order, all binary words of length n. (*Hint:* Exercise 15(a).)

17 The problem in Exercise 16 is to generate and list binary words in dictionary order. Here, the problem is to generate and list binary words in a different

order, one in which adjacent words differ in a single bit.* Because the kth word differs from its immediate predecessor in a single bit, to solve this problem it suffices to identify that bit. Here is a procedure for doing that: Every bit of the first word is zero. For $1 < k \leq 2^n$, the kth word is obtained from its predecessor by changing the dth bit, where $d-1$ is the highest power of 2 that exactly divides $k-1$.

(a) List the 16 binary words of length 4 in the order prescribed by this procedure. (*Hint:* As you go along, check to be sure that each newly listed word is different from all of its predecessors, and that it differs from its immediate predecessor in a single bit.)

(b) Show that word k differs from word $2^n - k + 1$ in a single bit, $1 \leq k \leq 2^n$.

(c) Show that the procedure described in this exercise generates 2^n different binary words of length n.

(d) Write an algorithm to implement the procedure described in the introduction to this exercise.

(e) Write an algorithm to list the 2^n subsets of $\{1, 2, \ldots, n\}$ in such a way that any two adjacent subsets on the list differ by just one element.

18 Assuming the keyword RND returns a pseudorandom† number from the interval (0, 1), the following subroutine will generate 1000 pseudorandom integers from the interval [0, 9]:

```
1. For I = 1 to 1000.
2. R(I) = ⌊10 × RND⌋.
3. Next I.
```

To the extent that RND simulates a true random-number generator, each integer in [0, 9] ought to occur with equal likelihood. Each time the subroutine is implemented, one would expect the number 9, e.g., to occur about 100 times.

(a) Write a computer program based on (an appropriate modification of) the subroutine to generate and output 50 pseudorandom integers between 0 and 9 (inclusive).

(b) Run your program from part (a) ten times (using ten different randomizing ''seeds'') and record the number of 9's that are produced in each run.

(c) Modify your program from part (a) to generate and print out 500 pseudorandom integers between 0 and 9 (inclusive) and, at the end, to output the number of 9's that were printed.

*A list in which each entry differs as little as possible from its predecessor is commonly called a ''Gray code''. Because such lists have nothing to do with binary codes, ''Gray list'' might be a better name for them.

†An algorithm to generate random numbers is something of an oxymoron. Truly random numbers are surprisingly difficult to obtain.

19 Assuming keyword RND returns a pseudorandom number, here is an algorithm to simulate the flipping of a single fair coin:

```
1. X = RND.
2. If X < 1/2, then write "H".
3. If X ≥ 1/2, then write "T".
```

(a) Write an algorithm to output 100 simulated flips of a fair coin.

(b) If you were to run a computer program that implements your algorithm from part (a), how many H's would you expect to see?

(c) Write a computer program to implement your algorithm from part (a), run it ten times (with ten different randomizing "seeds"), and record the total number of H's produced on each run.

(d) Write an algorithm to output 100 simulated flips of a fair coin and, at the end, output the total numbers of heads and tails.

(e) Write an algorithm to output 100 simulated flips of a fair coin and, at the end, output the (empirical) probability of heads.

20 If a fair coin is flipped 100 times, it would not be unusual to see a string of four or five heads in a row.

(a) Run your program from Exercise 19(c) ten times (using ten different randomizing "seeds") and record the longest string of consecutive H's and the longest string of consecutive T's for each run.

(b) Modify your algorithm/program from Exercise 19(a)/(c) so that it outputs the length of a longest string of consecutive H's and of a longest string of consecutive T's.

21 Suppose 12 fair coins are tossed into the air at once.

(a) Compute the probability of six heads and six tails.

(b) Write an algorithm to simulate 100 trials of tossing a dozen coins and to output the empirical probability that half the coins come up heads and half tails. (See the discussion of the keyword RND in the introduction to Exercise 18.)

22 Write an algorithm to simulate 100 flips of a biased coin, one in which heads occurs a third of the time. (*Hint:* See the introduction to Exercise 19.)

23 Write an algorithm to simulate 100 rolls of a fair die. (See the introduction to Exercise 18 for an explanation of the keyword RND.)

24 Assuming keyword RND returns a pseudorandom number, write an algorithm to simulate 1200 trials of rolling two (fair) dice

(a) and output the results.

(b) and output the empirical probability of rolling a (total of) 7.

25 Assuming keyword RND returns a pseudorandom number, write an algorithm to simulate 1200 trials of rolling a single (fair) dodecahedral die, and to output the results and the empirical probability of rolling a 7. (*Hint:* A dodecahedral die has twelve faces numbered 1–12.)

26 Assuming keyword RND returns a pseudorandom number, write an algorithm to simulate 1200 trials of rolling five (fair) dodecahedral dice and output the empirical probability of rolling a (sum of) 30.