Chapter 3

Combinatorics

3.1 Permutations

Many problems in probability theory require that we count the number of ways that a particular event can occur. For this, we study the topics of *permutations* and *combinations.* We consider permutations in this section and combinations in the next section.

Before discussing permutations, it is useful to introduce a general counting technique that will enable us to solve a variety of counting problems, including the problem of counting the number of possible permutations of n objects.

Counting Problems

Consider an experiment that takes place in several stages and is such that the number of outcomes m at the nth stage is independent of the outcomes of the previous stages. The number m may be different for different stages. We want to count the number of ways that the entire experiment can be carried out.

Example 3.1 You are eating at Émile's restaurant and the waiter informs you that you have (a) two choices for appetizers: soup or juice; (b) three for the main course: a meat, fish, or vegetable dish; and (c) two for dessert: ice cream or cake. How many possible choices do you have for your complete meal? We illustrate the possible meals by a tree diagram shown in Figure 3.1. Your menu is decided in three stages—at each stage the number of possible choices does not depend on what is chosen in the previous stages: two choices at the first stage, three at the second, and two at the third. From the tree diagram we see that the total number of choices is the product of the number of choices at each stage. In this examples we have $2 \cdot 3 \cdot 2 = 12$ possible menus. Our menu example is an example of the following general counting technique. □

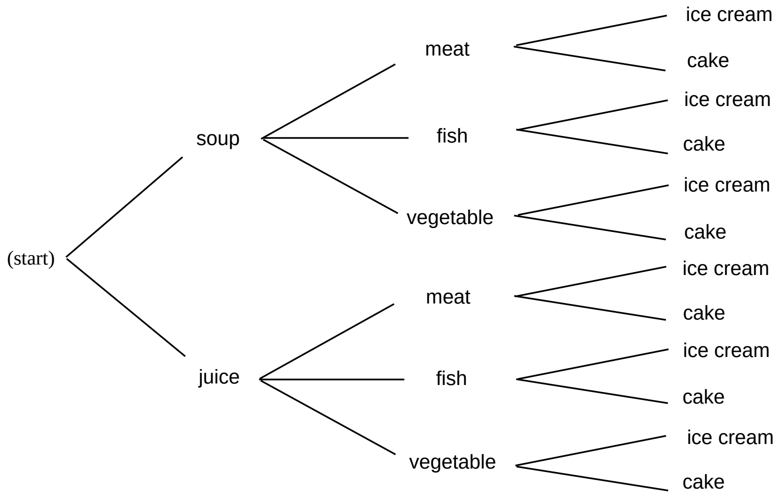

Figure 3.1: Tree for your menu.

A Counting Technique

A task is to be carried out in a sequence of r stages. There are n_1 ways to carry out the first stage; for each of these n_1 ways, there are n_2 ways to carry out the second stage; for each of these n_2 ways, there are n_3 ways to carry out the third stage, and so forth. Then the total number of ways in which the entire task can be accomplished is given by the product $N = n_1 \cdot n_2 \cdot \ldots \cdot n_r$.

Tree Diagrams

It will often be useful to use a tree diagram when studying probabilities of events relating to experiments that take place in stages and for which we are given the probabilities for the outcomes at each stage. For example, assume that the owner of Émile's restaurant has observed that 80 percent of his customers choose the soup for an appetizer and 20 percent choose juice. Of those who choose soup, 50 percent choose meat, 30 percent choose fish, and 20 percent choose the vegetable dish. Of those who choose juice for an appetizer, 30 percent choose meat, 40 percent choose fish, and 30 percent choose the vegetable dish. We can use this to estimate the probabilities at the first two stages as indicated on the tree diagram of Figure 3.2.

We choose for our sample space the set Ω of all possible paths $\omega = \omega_1$, ω_2, …, ω_6 through the tree. How should we assign our probability distribution? For example, what probability should we assign to the customer choosing soup and then the meat? If 8/10 of the customers choose soup and then 1/2 of these choose meat, a proportion $8/10 \cdot 1/2 = 4/10$ of the customers choose soup and then meat. This suggests choosing our probability distribution for each path through the tree to be the *product* of the probabilities at each of the stages along the path. This results in the probability measure for the sample points ω indicated in Figure 3.2. (Note that $m(\omega_1) + \cdots + m(\omega_6) = 1$.) From this we see, for example, that the probability

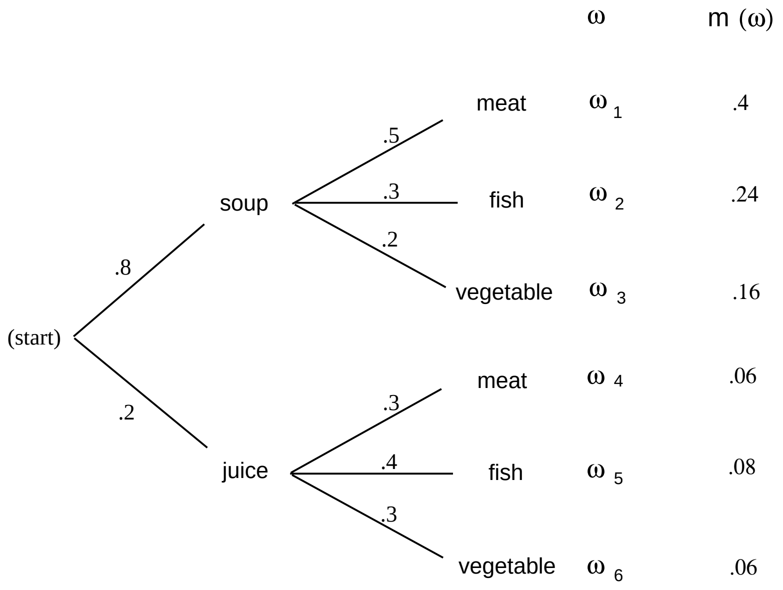

Figure 3.2: Two-stage probability assignment.

that a customer chooses meat is $m(\omega_1) + m(\omega_4) = .46$.

We shall say more about these tree measures when we discuss the concept of conditional probability in Chapter 4. We return now to more counting problems.

Example 3.2 We can show that there are at least two people in Columbus, Ohio, who have the same three initials. Assuming that each person has three initials, there are 26 possibilities for a person's first initial, 26 for the second, and 26 for the third. Therefore, there are $26^3 = 17{,}576$ possible sets of initials. This number is smaller than the number of people living in Columbus, Ohio; hence, there must be at least two people with the same three initials. □

We consider next the celebrated birthday problem—often used to show that naive intuition cannot always be trusted in probability.

Birthday Problem

Example 3.3 How many people do we need to have in a room to make it a favorable bet (probability of success greater than 1/2) that two people in the room will have the same birthday?

Since there are 365 possible birthdays, it is tempting to guess that we would need about 1/2 this number, or 183. You would surely win this bet. In fact, the number required for a favorable bet is only 23. To show this, we find the probability p_r that, in a room with r people, there is no duplication of birthdays; we will have a favorable bet if this probability is less than one half.

Number of people	Probability that all birthdays are different
20	.5885616
21	.5563117
22	.5243047
23	.4927028
24	.4616557
25	.4313003

Table 3.1: Birthday problem.

Assume that there are 365 possible birthdays for each person (we ignore leap years). Order the people from 1 to r. For a sample point ω, we choose a possible sequence of length r of birthdays each chosen as one of the 365 possible dates. There are 365 possibilities for the first element of the sequence, and for each of these choices there are 365 for the second, and so forth, making 365^r possible sequences of birthdays. We must find the number of these sequences that have no duplication of birthdays. For such a sequence, we can choose any of the 365 days for the first element, then any of the remaining 364 for the second, 363 for the third, and so forth, until we make r choices. For the rth choice, there will be $365 - r + 1$ possibilities. Hence, the total number of sequences with no duplications is

$$365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - r + 1)\ .$$

Thus, assuming that each sequence is equally likely,

$$p_r = \frac{365 \cdot 364 \cdot \ldots \cdot (365 - r + 1)}{365^r}\ .$$

We denote the product

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

by $(n)_r$ (read "n down r," or "n lower r"). Thus,

$$p_r = \frac{(365)_r}{(365)^r}\ .$$

The program **Birthday** carries out this computation and prints the probabilities for $r = 20$ to 25. Running this program, we get the results shown in Table 3.1. As we asserted above, the probability for no duplication changes from greater than one half to less than one half as we move from 22 to 23 people. To see how unlikely it is that we would lose our bet for larger numbers of people, we have run the program again, printing out values from $r = 10$ to $r = 100$ in steps of 10. We see that in a room of 40 people the odds already heavily favor a duplication, and in a room of 100 the odds are overwhelmingly in favor of a duplication. We have assumed that birthdays are equally likely to fall on any particular day. Statistical evidence suggests that this is not true. However, it is intuitively clear (but not easy to prove) that this makes it even more likely to have a duplication with a group of 23 people. (See Exercise 19 to find out what happens on planets with more or fewer than 365 days per year.) □

Number of people	Probability that all birthdays are different
10	.8830518
20	.5885616
30	.2936838
40	.1087682
50	.0296264
60	.0058773
70	.0008404
80	.0000857
90	.0000062
100	.0000003

Table 3.2: Birthday problem.

We now turn to the topic of permutations.

Permutations

Definition 3.1 Let A be any finite set. A *permutation of* A is a one-to-one mapping of A onto itself. □

To specify a particular permutation we list the elements of A and, under them, show where each element is sent by the one-to-one mapping. For example, if $A = \{a, b, c\}$ a possible permutation σ would be

$$\sigma = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix} .$$

By the permutation σ, a is sent to b, b is sent to c, and c is sent to a. The condition that the mapping be one-to-one means that no two elements of A are sent, by the mapping, into the same element of A.

We can put the elements of our set in some order and rename them 1, 2, …, n. Then, a typical permutation of the set $A = \{a_1, a_2, a_3, a_4\}$ can be written in the form

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} ,$$

indicating that a_1 went to a_2, a_2 to a_1, a_3 to a_4, and a_4 to a_3.

If we always choose the top row to be 1 2 3 4 then, to prescribe the permutation, we need only give the bottom row, with the understanding that this tells us where 1 goes, 2 goes, and so forth, under the mapping. When this is done, the permutation is often called a *rearrangement* of the n objects 1, 2, 3, …, n. For example, all possible permutations, or rearrangements, of the numbers $A = \{1, 2, 3\}$ are:

$$123,\ 132,\ 213,\ 231,\ 312,\ 321\ .$$

It is an easy matter to count the number of possible permutations of n objects. By our general counting principle, there are n ways to assign the first element, for

n	$n!$
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800

Table 3.3: Values of the factorial function.

each of these we have $n - 1$ ways to assign the second object, $n - 2$ for the third, and so forth. This proves the following theorem.

Theorem 3.1 The total number of permutations of a set A of n elements is given by $n \cdot (n \; - \; 1) \cdot (n - 2) \cdot \ldots \cdot 1$. □

It is sometimes helpful to consider orderings of subsets of a given set. This prompts the following definition.

Definition 3.2 Let A be an n-element set, and let k be an integer between 0 and n. Then a k-permutation of A is an ordered listing of a subset of A of size k. □

Using the same techniques as in the last theorem, the following result is easily proved.

Theorem 3.2 The total number of k-permutations of a set A of n elements is given by $n \cdot (n - 1) \cdot (n - 2) \cdot \ldots \cdot (n - k + 1)$. □

Factorials

The number given in Theorem 3.1 is called n *factorial,* and is denoted by $n!$. The expression $0!$ is defined to be 1 to make certain formulas come out simpler. The first few values of this function are shown in Table 3.3. The reader will note that this function grows very rapidly.

The expression $n!$ will enter into many of our calculations, and we shall need to have some estimate of its magnitude when n is large. It is clearly not practical to make exact calculations in this case. We shall instead use a result called *Stirling's formula.* Before stating this formula we need a definition.

n	$n!$	Approximation	Ratio
1	1	.922	1.084
2	2	1.919	1.042
3	6	5.836	1.028
4	24	23.506	1.021
5	120	118.019	1.016
6	720	710.078	1.013
7	5040	4980.396	1.011
8	40320	39902.395	1.010
9	362880	359536.873	1.009
10	3628800	3598696.619	1.008

Table 3.4: Stirling approximations to the factorial function.

Definition 3.3 Let a_n and b_n be two sequences of numbers. We say that a_n is *asymptotically equal to* b_n, and write $a_n \sim b_n$, if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1 \ .$$

□

Example 3.4 If $a_n = n + \sqrt{n}$ and $b_n = n$ then, since $a_n/b_n = 1 + 1/\sqrt{n}$ and this ratio tends to 1 as n tends to infinity, we have $a_n \sim b_n$. □

Theorem 3.3 (Stirling's Formula) The sequence $n!$ is asymptotically equal to

$$n^n e^{-n} \sqrt{2\pi n} \ .$$

□

The proof of Stirling's formula may be found in most analysis texts. Let us verify this approximation by using the computer. The program **StirlingApproximations** prints $n!$, the Stirling approximation, and, finally, the ratio of these two numbers. Sample output of this program is shown in Table 3.4. Note that, while the ratio of the numbers is getting closer to 1, the difference between the exact value and the approximation is increasing, and indeed, this difference will tend to infinity as n tends to infinity, even though the ratio tends to 1. (This was also true in our Example 3.4 where $n + \sqrt{n} \sim n$, but the difference is $\sqrt{n}$.)

Generating Random Permutations

We now consider the question of generating a random permutation of the integers between 1 and n. Consider the following experiment. We start with a deck of n cards, labelled 1 through n. We choose a random card out of the deck, note its label, and put the card aside. We repeat this process until all n cards have been chosen. It is clear that each permutation of the integers from 1 to n can occur as a sequence

Number of fixed points	Fraction of permutations		
	n = 10	n = 20	n = 30
0	.362	.370	.358
1	.368	.396	.358
2	.202	.164	.192
3	.052	.060	.070
4	.012	.008	.020
5	.004	.002	.002
Average number of fixed points	.996	.948	1.042

Table 3.5: Fixed point distributions.

of labels in this experiment, and that each sequence of labels is equally likely to occur. In our implementations of the computer algorithms, the above procedure is called **RandomPermutation**.

Fixed Points

There are many interesting problems that relate to properties of a permutation chosen at random from the set of all permutations of a given finite set. For example, since a permutation is a one-to-one mapping of the set onto itself, it is interesting to ask how many points are mapped onto themselves. We call such points *fixed points* of the mapping.

Let $p_k(n)$ be the probability that a random permutation of the set $\{1, 2, \ldots, n\}$ has exactly k fixed points. We will attempt to learn something about these probabilities using simulation. The program **FixedPoints** uses the procedure **RandomPermutation** to generate random permutations and count fixed points. The program prints the proportion of times that there are k fixed points as well as the average number of fixed points. The results of this program for 500 simulations for the cases $n = 10$, 20, and 30 are shown in Table 3.5. Notice the rather surprising fact that our estimates for the probabilities do not seem to depend very heavily on the number of elements in the permutation. For example, the probability that there are no fixed points, when $n = 10,\ 20$, or 30 is estimated to be between .35 and .37. We shall see later (see Example 3.12) that for $n \geq 10$ the exact probabilities $p_n(0)$ are, to six decimal place accuracy, equal to $1/e \approx .367879$. Thus, for all practical purposes, after $n = 10$ the probability that a random permutation of the set $\{1, 2, \ldots, n\}$ does not depend upon n. These simulations also suggest that the average number of fixed points is close to 1. It can be shown (see Example 6.8) that the average is exactly equal to 1 for all n.

More picturesque versions of the fixed-point problem are: You have arranged the books on your book shelf in alphabetical order by author and they get returned to your shelf at random; what is the probability that exactly k of the books end up in their correct position? (The library problem.) In a restaurant n hats are checked and they are hopelessly scrambled; what is the probability that no one gets his own hat back? (The hat check problem.) In the Historical Remarks at the end of this section, we give one method for solving the hat check problem exactly. Another

Date	Snowfall in inches
1974	75
1975	88
1976	72
1977	110
1978	85
1979	30
1980	55
1981	86
1982	51
1983	64

Table 3.6: Snowfall in Hanover.

Year	1	2	3	4	5	6	7	8	9	10
Ranking	6	9	5	10	7	1	3	8	2	4

Table 3.7: Ranking of total snowfall.

method is given in Example 3.12.

Records

Here is another interesting probability problem that involves permutations. Estimates for the amount of measured snow in inches in Hanover, New Hampshire, in the ten years from 1974 to 1983 are shown in Table 3.6. Suppose we have started keeping records in 1974. Then our first year's snowfall could be considered a record snowfall starting from this year. A new record was established in 1975; the next record was established in 1977, and there were no new records established after this year. Thus, in this ten-year period, there were three records established: 1974, 1975, and 1977. The question that we ask is: How many records should we expect to be established in such a ten-year period? We can count the number of records in terms of a permutation as follows: We number the years from 1 to 10. The actual amounts of snowfall are not important but their relative sizes are. We can, therefore, change the numbers measuring snowfalls to numbers 1 to 10 by replacing the smallest number by 1, the next smallest by 2, and so forth. (We assume that there are no ties.) For our example, we obtain the data shown in Table 3.7.

This gives us a permutation of the numbers from 1 to 10 and, from this permutation, we can read off the records; they are in years 1, 2, and 4. Thus we can define records for a permutation as follows:

Definition 3.4 Let σ be a permutation of the set $\{1, 2, \ldots, n\}$. Then i is a *record* of σ if either $i = 1$ or $\sigma(j) < \sigma(i)$ for every $j = 1, \ldots, i - 1$. □

Now if we regard all rankings of snowfalls over an n-year period to be equally likely (and allow no ties), we can estimate the probability that there will be k records in n years as well as the average number of records by simulation.

We have written a program **Records** that counts the number of records in randomly chosen permutations. We have run this program for the cases $n = 10, 20, 30$. For $n = 10$ the average number of records is 2.968, for 20 it is 3.656, and for 30 it is 3.960. We see now that the averages increase, but very slowly. We shall see later (see Example 6.11) that the average number is approximately $\log n$. Since $\log 10 = 2.3$, $\log 20 = 3$, and $\log 30 = 3.4$, this is consistent with the results of our simulations.

As remarked earlier, we shall be able to obtain formulas for exact results of certain problems of the above type. However, only minor changes in the problem make this impossible. The power of simulation is that minor changes in a problem do not make the simulation much more difficult. (See Exercise 20 for an interesting variation of the hat check problem.)

List of Permutations

Another method to solve problems that is not sensitive to small changes in the problem is to have the computer simply list all possible permutations and count the fraction that have the desired property. The program **AllPermutations** produces a list of all of the permutations of n. When we try running this program, we run into a limitation on the use of the computer. The number of permutations of n increases so rapidly that even to list all permutations of 20 objects is impractical.

Historical Remarks

Our basic counting principle stated that if you can do one thing in r ways and for each of these another thing in s ways, then you can do the pair in rs ways. This is such a self-evident result that you might expect that it occurred very early in mathematics. N. L. Biggs suggests that we might trace an example of this principle as follows: First, he relates a popular nursery rhyme dating back to at least 1730:

> As I was going to St. Ives,
> I met a man with seven wives,
> Each wife had seven sacks,
> Each sack had seven cats,
> Each cat had seven kits.
> Kits, cats, sacks and wives,
> How many were going to St. Ives?

(You need our principle only if you are not clever enough to realize that you are supposed to answer *one,* since only the narrator is going to St. Ives; the others are going in the other direction!)

He also gives a problem appearing on one of the oldest surviving mathematical manuscripts of about 1650 B.C., roughly translated as:

Houses	7
Cats	49
Mice	343
Wheat	2401
Hekat	16807
	19607

The following interpretation has been suggested: there are seven houses, each with seven cats; each cat kills seven mice; each mouse would have eaten seven heads of wheat, each of which would have produced seven hekat measures of grain. With this interpretation, the table answers the question of how many hekat measures were saved by the cats' actions. It is not clear why the writer of the table wanted to add the numbers together.[1]

One of the earliest uses of factorials occurred in Euclid's proof that there are infinitely many prime numbers. Euclid argued that there must be a prime number between n and $n! + 1$ as follows: $n!$ and $n! + 1$ cannot have common factors. Either $n! + 1$ is prime or it has a proper factor. In the latter case, this factor cannot divide $n!$ and hence must be between n and $n! + 1$. If this factor is not prime, then it has a factor that, by the same argument, must be bigger than n. In this way, we eventually reach a prime bigger than n, and this holds for all n.

The "$n!$" rule for the number of permutations seems to have occurred first in India. Examples have been found as early as 300 B.C., and by the eleventh century the general formula seems to have been well known in India and then in the Arab countries.

The *hat check problem* is found in an early probability book written by de Montmort and first printed in 1708.[2] It appears in the form of a game called *Treize.* In a simplified version of this game considered by de Montmort one turns over cards numbered 1 to 13, calling out 1, 2, . . . , 13 as the cards are examined. De Montmort asked for the probability that no card that is turned up agrees with the number called out.

This probability is the same as the probability that a random permutation of 13 elements has no fixed point. De Montmort solved this problem by the use of a recursion relation as follows: let w_n be the number of permutations of n elements with no fixed point (such permutations are called *derangements*). Then $w_1 = 0$ and $w_2 = 1$.

Now assume that $n \geq 3$ and choose a derangement of the integers between 1 and n. Let k be the integer in the first position in this derangement. By the definition of derangement, we have $k \neq 1$. There are two possibilities of interest concerning the position of 1 in the derangement: either 1 is in the kth position or it is elsewhere. In the first case, the $n - 2$ remaining integers can be positioned in w_{n-2} ways without resulting in any fixed points. In the second case, we consider the set of integers $\{1, 2, \ldots, k - 1, k + 1, \ldots, n\}$. The numbers in this set must occupy the positions $\{2, 3, \ldots, n\}$ so that none of the numbers other than 1 in this set are fixed, and

[1] N. L. Biggs, "The Roots of Combinatorics," *Historia Mathematica,* vol. 6 (1979), pp. 109–136.

[2] P. R. de Montmort, *Essay d'Analyse sur des Jeux de Hazard,* 2d ed. (Paris: Quillau, 1713).

also so that 1 is not in position k. The number of ways of achieving this kind of arrangement is just w_{n-1}. Since there are $n-1$ possible values of k, we see that

$$w_n = (n-1)w_{n-1} + (n-1)w_{n-2}$$

for $n \geq 3$. One might conjecture from this last equation that the sequence $\{w_n\}$ grows like the sequence $\{n!\}$.

In fact, it is easy to prove by induction that

$$w_n = nw_{n-1} + (-1)^n \ .$$

Then $p_i = w_i/i!$ satisfies

$$p_i - p_{i-1} = \frac{(-1)^i}{i!} \ .$$

If we sum from $i = 2$ to n, and use the fact that $p_1 = 0$, we obtain

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

This agrees with the first $n+1$ terms of the expansion for e^x for $x = -1$ and hence for large n is approximately $e^{-1} \approx .368$. David remarks that this was possibly the first use of the exponential function in probability.[3] We shall see another way to derive de Montmort's result in the next section, using a method known as the Inclusion-Exclusion method.

Recently, a related problem appeared in a column of Marilyn vos Savant.[4] Charles Price wrote to ask about his experience playing a certain form of solitaire, sometimes called "frustration solitaire." In this particular game, a deck of cards is shuffled, and then dealt out, one card at a time. As the cards are being dealt, the player counts from 1 to 13, and then starts again at 1. (Thus, each number is counted four times.) If a number that is being counted coincides with the rank of the card that is being turned up, then the player loses the game. Price found that he he rarely won and wondered how often he should win. Vos Savant remarked that the expected number of matches is 4 so it should be difficult to win the game.

Finding the chance of winning is a harder problem than the one that de Montmort solved because, when one goes through the entire deck, there are different patterns for the matches that might occur. For example matches may occur for two cards of the same rank, say two aces, or for two different ranks, say a two and a three.

A discussion of this problem can be found in Riordan.[5] In this book, it is shown that as $n \to \infty$, the probability of no matches tends to $1/e^4$.

The original game of Treize is more difficult to analyze than frustration solitaire. The game of Treize is played as follows. One person is chosen as dealer and the others are players. Each player, other than the dealer, puts up a stake. The dealer shuffles the cards and turns them up one at a time calling out, "Ace, two, three,...,

[3] F. N. David, *Games, Gods and Gambling* (London: Griffin, 1962), p. 146.

[4] M. vos Savant, Ask Marilyn, *Parade Magazine, Boston Globe*, 21 August 1994.

[5] J. Riordan, *An Introduction to Combinatorial Analysis,* (New York: John Wiley & Sons, 1958).

king," just as in frustration solitaire. If the dealer goes through the 13 cards without a match he pays the players an amount equal to their stake, and the deal passes to someone else. If there is a match the dealer collects the players' stakes; the players put up new stakes, and the dealer continues through the deck, calling out, "Ace, two, three, ..." If the dealer runs out of cards he reshuffles and continues the count where he left off. He continues until there is a run of 13 without a match and then a new dealer is chosen.

The question at this point is how much money can the dealer expect to win from each player. De Montmort found that if each player puts up a stake of 1, say, then the dealer will win approximately .801 from each player.

Peter Doyle calculated the exact amount that the dealer can expect to win. The answer is:

26516072156010218582227607912734182784642120482136091446715371962089931
52311343541724554334912870541440299239251607694113500080775917818512013
82176876653563173852874555859367254632009477403727395572807459384342747
87664965076063990538261189388143513547366316017004945507201764278828306
60117107953633142734382477922709835281753299035988581413688367655833113
24476153310720627474169719301806649152698704084383914217907906954976036
28528211590140316202120601549126920880824913325553882692055427830810368
5781886120875824880068097864043811858283487754256095550662878927123048
26997601700116233592793308297533642193505074540268925683193887821301442
70519791882/
33036929133582592220117220713156071114975101149831063364072138969878007
99647204708825303387525892236581323015628005621143427290625658974433971
65719454122908007086289841306087561302818991167357863623756067184986491
35353553622197448890223267101158801016285931351979294387223277033396967
79797069933475802423676949873661605184031477561560393380257070970711959
69641268242455013319879747054693517809383750593488858698672364846950539
88868628582609905586271001318150621134407056983214740221851567706672080
94586589378459432799868706334161812988630496327287254818458879353024498
00322425586446741048147720934108061350613503856973048971213063937040515
59533731591.

This is .803 to 3 decimal places. A description of the algorithm used to find this answer can be found on his Web page.[6] A discussion of this problem and other problems can be found in Doyle et al.[7]

The *birthday problem* does not seem to have a very old history. Problems of this type were first discussed by von Mises.[8] It was made popular in the 1950s by Feller's book.[9]

[6] P. Doyle, "Solution to Montmort's Probleme du Treize," http://math.ucsd.edu/~doyle/.

[7] P. Doyle, C. Grinstead, and J. Snell, "Frustration Solitaire," *UMAP Journal*, vol. 16, no. 2 (1995), pp. 137-145.

[8] R. von Mises, "Über Aufteilungs- und Besetzungs-Wahrscheinlichkeiten," *Revue de la Faculté des Sciences de l'Université d'Istanbul, N. S.* vol. 4 (1938-39), pp. 145-163.

[9] W. Feller, *Introduction to Probability Theory and Its Applications,* vol. 1, 3rd ed. (New York:

Stirling presented his formula

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

in his work *Methodus Differentialis* published in 1730.[10] This approximation was used by de Moivre in establishing his celebrated central limit theorem that we will study in Chapter 9. De Moivre himself had independently established this approximation, but without identifying the constant π. Having established the approximation

$$\frac{2B}{\sqrt{n}}$$

for the central term of the binomial distribution, where the constant B was determined by an infinite series, de Moivre writes:

> . . . my worthy and learned Friend, Mr. James Stirling, who had applied himself after me to that inquiry, found that the Quantity B did denote the Square-root of the Circumference of a Circle whose Radius is Unity, so that if that Circumference be called c the Ratio of the middle Term to the Sum of all Terms will be expressed by $2/\sqrt{nc}\,\ldots.$[11]

Exercises

1 Four people are to be arranged in a row to have their picture taken. In how many ways can this be done?

2 An automobile manufacturer has four colors available for automobile exteriors and three for interiors. How many different color combinations can he produce?

3 In a digital computer, a *bit* is one of the integers {0,1}, and a *word* is any string of 32 bits. How many different words are possible?

4 What is the probability that at least 2 of the presidents of the United States have died on the same day of the year? If you bet this has happened, would you win your bet?

5 There are three different routes connecting city A to city B. How many ways can a round trip be made from A to B and back? How many ways if it is desired to take a different route on the way back?

6 In arranging people around a circular table, we take into account their seats relative to each other, not the actual position of any one person. Show that n people can be arranged around a circular table in $(n-1)!$ ways.

John Wiley & Sons, 1968).

[10] J. Stirling, *Methodus Differentialis,* (London: Bowyer, 1730).

[11] A. de Moivre, *The Doctrine of Chances,* 3rd ed. (London: Millar, 1756).

7 Five people get on an elevator that stops at five floors. Assuming that each has an equal probability of going to any one floor, find the probability that they all get off at different floors.

8 A finite set Ω has n elements. Show that if we count the empty set and Ω as subsets, there are 2^n subsets of Ω.

9 A more refined inequality for approximating $n!$ is given by

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n)} .$$

Write a computer program to illustrate this inequality for $n = 1$ to 9.

10 A deck of ordinary cards is shuffled and 13 cards are dealt. What is the probability that the last card dealt is an ace?

11 There are n applicants for the director of computing. The applicants are interviewed independently by each member of the three-person search committee and ranked from 1 to n. A candidate will be hired if he or she is ranked first by at least two of the three interviewers. Find the probability that a candidate will be accepted if the members of the committee really have no ability at all to judge the candidates and just rank the candidates randomly. In particular, compare this probability for the case of three candidates and the case of ten candidates.

12 A symphony orchestra has in its repertoire 30 Haydn symphonies, 15 modern works, and 9 Beethoven symphonies. Its program always consists of a Haydn symphony followed by a modern work, and then a Beethoven symphony.

(a) How many different programs can it play?

(b) How many different programs are there if the three pieces can be played in any order?

(c) How many different three-piece programs are there if more than one piece from the same category can be played and they can be played in any order?

13 A certain state has license plates showing three numbers and three letters. How many different license plates are possible

(a) if the numbers must come before the letters?

(b) if there is no restriction on where the letters and numbers appear?

14 The door on the computer center has a lock which has five buttons numbered from 1 to 5. The combination of numbers that opens the lock is a sequence of five numbers and is reset every week.

(a) How many combinations are possible if every button must be used once?

(b) Assume that the lock can also have combinations that require you to push two buttons simultaneously and then the other three one at a time. How many more combinations does this permit?

15 A computing center has 3 processors that receive n jobs, with the jobs assigned to the processors purely at random so that all of the 3^n possible assignments are equally likely. Find the probability that exactly one processor has no jobs.

16 Prove that at least two people in Atlanta, Georgia, have the same initials, assuming no one has more than four initials.

17 Find a formula for the probability that among a set of n people, at least two have their birthdays in the same month of the year (assuming the months are equally likely for birthdays).

18 Consider the problem of finding the probability of more than one coincidence of birthdays in a group of n people. These include, for example, three people with the same birthday, or two pairs of people with the same birthday, or larger coincidences. Show how you could compute this probability, and write a computer program to carry out this computation. Use your program to find the smallest number of people for which it would be a favorable bet that there would be more than one coincidence of birthdays.

***19** Suppose that on planet Zorg a year has n days, and that the lifeforms there are equally likely to have hatched on any day of the year. We would like to estimate d, which is the minimum number of lifeforms needed so that the probability of at least two sharing a birthday exceeds 1/2.

(a) In Example 3.3, it was shown that in a set of d lifeforms, the probability that no two life forms share a birthday is

$$\frac{(n)_d}{n^d}\ ,$$

where $(n)_d = (n)(n-1)\cdots(n-d+1)$. Thus, we would like to set this equal to 1/2 and solve for d.

(b) Using Stirling's Formula, show that

$$\frac{(n)_d}{n^d} \sim \left(1 + \frac{d}{n-d}\right)^{n-d+1/2} e^{-d}\ .$$

(c) Now take the logarithm of the right-hand expression, and use the fact that for small values of x, we have

$$\log(1+x) \sim x - \frac{x^2}{2}\ .$$

(We are implicitly using the fact that d is of smaller order of magnitude than n. We will also use this fact in part (d).)

(d) Set the expression found in part (c) equal to $-\log(2)$, and solve for d as a function of n, thereby showing that

$$d \sim \sqrt{2(\log 2)\,n}\ .$$

Hint: If all three summands in the expression found in part (b) are used, one obtains a cubic equation in d. If the smallest of the three terms is thrown away, one obtains a quadratic equation in d.

(e) Use a computer to calculate the exact values of d for various values of n. Compare these values with the approximate values obtained by using the answer to part d).

20 At a mathematical conference, ten participants are randomly seated around a circular table for meals. Using simulation, estimate the probability that no two people sit next to each other at both lunch and dinner. Can you make an intelligent conjecture for the case of n participants when n is large?

21 Modify the program **AllPermutations** to count the number of permutations of n objects that have exactly j fixed points for $j = 0,\ 1,\ 2,\ \ldots,\ n$. Run your program for $n = 2$ to 6. Make a conjecture for the relation between the number that have 0 fixed points and the number that have exactly 1 fixed point. A proof of the correct conjecture can be found in Wilf.[12]

22 Mr. Wimply Dimple, one of London's most prestigious watch makers, has come to Sherlock Holmes in a panic, having discovered that someone has been producing and selling crude counterfeits of his best selling watch. The 16 counterfeits so far discovered bear stamped numbers, all of which fall between 1 and 56, and Dimple is anxious to know the extent of the forger's work. All present agree that it seems reasonable to assume that the counterfeits thus far produced bear consecutive numbers from 1 to whatever the total number is.

"Chin up, Dimple," opines Dr. Watson. "I shouldn't worry overly much if I were you; the Maximum Likelihood Principle, which estimates the total number as precisely that which gives the highest probability for the series of numbers found, suggests that we guess 56 itself as the total. Thus, your forgers are not a big operation, and we shall have them safely behind bars before your business suffers significantly."

"Stuff, nonsense, and bother your fancy principles, Watson," counters Holmes. "Anyone can see that, of course, there must be quite a few more than 56 watches—why the odds of our having discovered precisely the highest numbered watch made are laughably negligible. A much better guess would be *twice* 56."

(a) Show that Watson is correct that the Maximum Likelihood Principle gives 56.

[12] H. S. Wilf, "A Bijection in the Theory of Derangements," *Mathematics Magazine,* vol. 57, no. 1 (1984), pp. 37–40.

(b) Write a computer program to compare Holmes's and Watson's guessing strategies as follows: fix a total N and choose 16 integers randomly between 1 and N. Let m denote the largest of these. Then Watson's guess for N is m, while Holmes's is $2m$. See which of these is closer to N. Repeat this experiment (with N still fixed) a hundred or more times, and determine the proportion of times that each comes closer. Whose seems to be the better strategy?

23 Barbara Smith is interviewing candidates to be her secretary. As she interviews the candidates, she can determine the relative rank of the candidates but not the true rank. Thus, if there are six candidates and their true rank is 6, 1, 4, 2, 3, 5, (where 1 is best) then after she had interviewed the first three candidates she would rank them 3, 1, 2. As she interviews each candidate, she must either accept or reject the candidate. If she does not accept the candidate after the interview, the candidate is lost to her. She wants to decide on a strategy for deciding when to stop and accept a candidate that will maximize the probability of getting the best candidate. Assume that there are n candidates and they arrive in a random rank order.

(a) What is the probability that Barbara gets the best candidate if she interviews all of the candidates? What is it if she chooses the first candidate?

(b) Assume that Barbara decides to interview the first half of the candidates and then continue interviewing until getting a candidate better than any candidate seen so far. Show that she has a better than 25 percent chance of ending up with the best candidate.

24 For the task described in Exercise 23, it can be shown[13] that the best strategy is to pass over the first $k - 1$ candidates where k is the smallest integer for which

$$\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{n-1} \leq 1 \ .$$

Using this strategy the probability of getting the best candidate is approximately $1/e = .368$. Write a program to simulate Barbara Smith's interviewing if she uses this optimal strategy, using $n = 10$, and see if you can verify that the probability of success is approximately $1/e$.

3.2 Combinations

Having mastered permutations, we now consider combinations. Let U be a set with n elements; we want to count the number of distinct subsets of the set U that have exactly j elements. The empty set and the set U are considered to be subsets of U. The empty set is usually denoted by ϕ.

[13] E. B. Dynkin and A. A. Yushkevich, *Markov Processes: Theorems and Problems,* trans. J. S. Wood (New York: Plenum, 1969).

Example 3.5 Let $U = \{a, b, c\}$. The subsets of U are

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

□

Binomial Coefficients

The number of distinct subsets with j elements that can be chosen from a set with n elements is denoted by $\binom{n}{j}$, and is pronounced "n choose j." The number $\binom{n}{j}$ is called a *binomial coefficient.* This terminology comes from an application to algebra which will be discussed later in this section.

In the above example, there is one subset with no elements, three subsets with exactly 1 element, three subsets with exactly 2 elements, and one subset with exactly 3 elements. Thus, $\binom{3}{0} = 1$, $\binom{3}{1} = 3$, $\binom{3}{2} = 3$, and $\binom{3}{3} = 1$. Note that there are $2^3 = 8$ subsets in all. (We have already seen that a set with n elements has 2^n subsets; see Exercise 3.1.8.) It follows that

$$\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 2^3 = 8\ ,$$

$$\binom{n}{0} = \binom{n}{n} = 1\ .$$

Assume that $n > 0$. Then, since there is only one way to choose a set with no elements and only one way to choose a set with n elements, the remaining values of $\binom{n}{j}$ are determined by the following *recurrence relation*:

Theorem 3.4 For integers n and j, with $0 < j < n$, the binomial coefficients satisfy:

$$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}\ . \tag{3.1}$$

Proof. We wish to choose a subset of j elements. Choose an element u of U. Assume first that we do not want u in the subset. Then we must choose the j elements from a set of $n - 1$ elements; this can be done in $\binom{n-1}{j}$ ways. On the other hand, assume that we do want u in the subset. Then we must choose the other $j - 1$ elements from the remaining $n - 1$ elements of U; this can be done in $\binom{n-1}{j-1}$ ways. Since u is either in our subset or not, the number of ways that we can choose a subset of j elements is the sum of the number of subsets of j elements which have u as a member and the number which do not—this is what Equation 3.1 states. □

The binomial coefficient $\binom{n}{j}$ is defined to be 0, if $j < 0$ or if $j > n$. With this definition, the restrictions on j in Theorem 3.4 are unnecessary.

	j = 0	1	2	3	4	5	6	7	8	9	10
n = 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			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

Figure 3.3: Pascal's triangle.

Pascal's Triangle

The relation 3.1, together with the knowledge that

$$\binom{n}{0} = \binom{n}{n} = 1 \ ,$$

determines completely the numbers $\binom{n}{j}$. We can use these relations to determine the famous *triangle of Pascal,* which exhibits all these numbers in matrix form (see Figure 3.3).

The nth row of this triangle has the entries $\binom{n}{0}$, $\binom{n}{1}$,…, $\binom{n}{n}$. We know that the first and last of these numbers are 1. The remaining numbers are determined by the recurrence relation Equation 3.1; that is, the entry $\binom{n}{j}$ for $0 < j < n$ in the nth row of Pascal's triangle is the *sum* of the entry immediately above and the one immediately to its left in the $(n-1)$st row. For example, $\binom{5}{2} = 6 + 4 = 10$.

This algorithm for constructing Pascal's triangle can be used to write a computer program to compute the binomial coefficients. You are asked to do this in Exercise 4.

While Pascal's triangle provides a way to construct recursively the binomial coefficients, it is also possible to give a formula for $\binom{n}{j}$.

Theorem 3.5 The binomial coefficients are given by the formula

$$\binom{n}{j} = \frac{(n)_j}{j!} \ . \tag{3.2}$$

Proof. Each subset of size j of a set of size n can be ordered in $j!$ ways. Each of these orderings is a j-permutation of the set of size n. The number of j-permutations is $(n)_j$, so the number of subsets of size j is

$$\frac{(n)_j}{j!} \ .$$

This completes the proof. □

The above formula can be rewritten in the form

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}\ .$$

This immediately shows that

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

When using Equation 3.2 in the calculation of $\binom{n}{j}$, if one alternates the multiplications and divisions, then all of the intermediate values in the calculation are integers. Furthermore, none of these intermediate values exceed the final value. (See Exercise 40.)

Another point that should be made concerning Equation 3.2 is that if it is used to *define* the binomial coefficients, then it is no longer necessary to require n to be a positive integer. The variable j must still be a non-negative integer under this definition. This idea is useful when extending the Binomial Theorem to general exponents. (The Binomial Theorem for non-negative integer exponents is given below as Theorem 3.7.)

Poker Hands

Example 3.6 Poker players sometimes wonder why a *four of a kind* beats a *full house.* A poker hand is a random subset of 5 elements from a deck of 52 cards. A hand has four of a kind if it has four cards with the same value—for example, four sixes or four kings. It is a full house if it has three of one value and two of a second—for example, three twos and two queens. Let us see which hand is more likely. How many hands have four of a kind? There are 13 ways that we can specify the value for the four cards. For each of these, there are 48 possibilities for the fifth card. Thus, the number of four-of-a-kind hands is $13 \cdot 48 = 624$. Since the total number of possible hands is $\binom{52}{5} = 2598960$, the probability of a hand with four of a kind is $624/2598960 = .00024$.

Now consider the case of a full house; how many such hands are there? There are 13 choices for the value which occurs three times; for each of these there are $\binom{4}{3} = 4$ choices for the particular three cards of this value that are in the hand. Having picked these three cards, there are 12 possibilities for the value which occurs twice; for each of these there are $\binom{4}{2} = 6$ possibilities for the particular pair of this value. Thus, the number of full houses is $13 \cdot 4 \cdot 12 \cdot 6 = 3744$, and the probability of obtaining a hand with a full house is $3744/2598960 = .0014$. Thus, while both types of hands are unlikely, you are six times more likely to obtain a full house than four of a kind. □

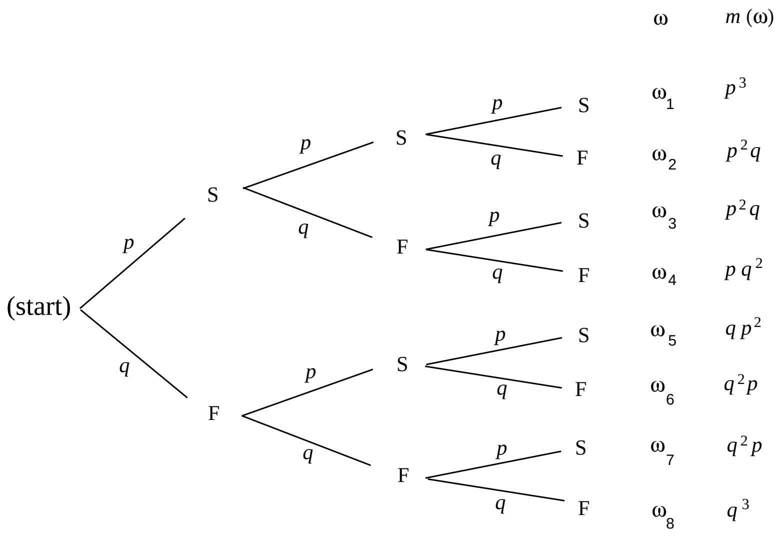

Figure 3.4: Tree diagram of three Bernoulli trials.

Bernoulli Trials

Our principal use of the binomial coefficients will occur in the study of one of the important chance processes called *Bernoulli trials.*

Definition 3.5 A *Bernoulli trials process* is a sequence of n chance experiments such that

1. Each experiment has two possible outcomes, which we may call *success* and *failure.*

2. The probability p of success on each experiment is the same for each experiment, and this probability is not affected by any knowledge of previous outcomes. The probability q of failure is given by $q = 1 - p$.

□

Example 3.7 The following are Bernoulli trials processes:

1. A coin is tossed ten times. The two possible outcomes are heads and tails. The probability of heads on any one toss is 1/2.

2. An opinion poll is carried out by asking 1000 people, randomly chosen from the population, if they favor the Equal Rights Amendment—the two outcomes being yes and no. The probability p of a yes answer (i.e., a success) indicates the proportion of people in the entire population that favor this amendment.

3. A gambler makes a sequence of 1-dollar bets, betting each time on black at roulette at Las Vegas. Here a success is winning 1 dollar and a failure is losing

1 dollar. Since in American roulette the gambler wins if the ball stops on one of 18 out of 38 positions and loses otherwise, the probability of winning is $p = 18/38 = .474$.

□

To analyze a Bernoulli trials process, we choose as our sample space a binary tree and assign a probability measure to the paths in this tree. Suppose, for example, that we have three Bernoulli trials. The possible outcomes are indicated in the tree diagram shown in Figure 3.4. We define X to be the random variable which represents the outcome of the process, i.e., an ordered triple of S's and F's. The probabilities assigned to the branches of the tree represent the probability for each individual trial. Let the outcome of the ith trial be denoted by the random variable X_i, with distribution function m_i. Since we have assumed that outcomes on any one trial do not affect those on another, we assign the same probabilities at each level of the tree. An outcome ω for the entire experiment will be a path through the tree. For example, ω_3 represents the outcomes SFS. Our frequency interpretation of probability would lead us to expect a fraction p of successes on the first experiment; of these, a fraction q of failures on the second; and, of these, a fraction p of successes on the third experiment. This suggests assigning probability pqp to the outcome ω_3. More generally, we assign a distribution function $m(\omega)$ for paths ω by defining $m(\omega)$ to be the product of the branch probabilities along the path ω. Thus, the probability that the three events S on the first trial, F on the second trial, and S on the third trial occur is the product of the probabilities for the individual events. We shall see in the next chapter that this means that the events involved are *independent* in the sense that the knowledge of one event does not affect our prediction for the occurrences of the other events.

Binomial Probabilities

We shall be particularly interested in the probability that in n Bernoulli trials there are exactly j successes. We denote this probability by $b(n, p, j)$. Let us calculate the particular value $b(3, p, 2)$ from our tree measure. We see that there are three paths which have exactly two successes and one failure, namely ω_2, ω_3, and ω_5. Each of these paths has the same probability p^2q. Thus $b(3, p, 2) = 3p^2q$. Considering all possible numbers of successes we have

$$\begin{aligned} b(3,p,0) &= q^3 \ , \\ b(3,p,1) &= 3pq^2 \ , \\ b(3,p,2) &= 3p^2q \ , \\ b(3,p,3) &= p^3 \ . \end{aligned}$$

We can, in the same manner, carry out a tree measure for n experiments and determine $b(n, p, j)$ for the general case of n Bernoulli trials.

Theorem 3.6 Given n Bernoulli trials with probability p of success on each experiment, the probability of exactly j successes is

$$b(n,p,j) = \binom{n}{j} p^j q^{n-j}$$

where $q = 1 - p$.

Proof. We construct a tree measure as described above. We want to find the sum of the probabilities for all paths which have exactly j successes and $n - j$ failures. Each such path is assigned a probability $p^j q^{n-j}$. How many such paths are there? To specify a path, we have to pick, from the n possible trials, a subset of j to be successes, with the remaining $n - j$ outcomes being failures. We can do this in $\binom{n}{j}$ ways. Thus the sum of the probabilities is

$$b(n,p,j) = \binom{n}{j} p^j q^{n-j} \ .$$

□

Example 3.8 A fair coin is tossed six times. What is the probability that exactly three heads turn up? The answer is

$$b(6,.5,3) = \binom{6}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3 = 20 \cdot \frac{1}{64} = .3125 \ .$$

□

Example 3.9 A die is rolled four times. What is the probability that we obtain exactly one 6? We treat this as Bernoulli trials with *success* = "rolling a 6" and *failure* = "rolling some number other than a 6." Then $p = 1/6$, and the probability of exactly one success in four trials is

$$b(4,1/6,1) = \binom{4}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^3 = .386 \ .$$

□

To compute binomial probabilities using the computer, multiply the function choose(n,k) by $p^k q^{n-k}$. The program **BinomialProbabilities** prints out the binomial probabilities $b(n,p,k)$ for k between *kmin* and *kmax*, and the sum of these probabilities. We have run this program for $n = 100$, $p = 1/2$, $kmin = 45$, and $kmax = 55$; the output is shown in Table 3.8. Note that the individual probabilities are quite small. The probability of exactly 50 heads in 100 tosses of a coin is about .08. Our intuition tells us that this is the most likely outcome, which is correct; but, all the same, it is not a very likely outcome.

k	$b(n, p, k)$
45	.0485
46	.0580
47	.0666
48	.0735
49	.0780
50	.0796
51	.0780
52	.0735
53	.0666
54	.0580
55	.0485

Table 3.8: Binomial probabilities for $n = 100,\ p = 1/2$.

Binomial Distributions

Definition 3.6 Let n be a positive integer, and let p be a real number between 0 and 1. Let B be the random variable which counts the number of successes in a Bernoulli trials process with parameters n and p. Then the distribution $b(n, p, k)$ of B is called the *binomial distribution*. □

We can get a better idea about the binomial distribution by graphing this distribution for different values of n and p (see Figure 3.5). The plots in this figure were generated using the program **BinomialPlot**.

We have run this program for $p = .5$ and $p = .3$. Note that even for $p = .3$ the graphs are quite symmetric. We shall have an explanation for this in Chapter 9. We also note that the highest probability occurs around the value np, but that these highest probabilities get smaller as n increases. We shall see in Chapter 6 that np is the *mean* or *expected* value of the binomial distribution $b(n, p, k)$.

The following example gives a nice way to see the binomial distribution, when $p = 1/2$.

Example 3.10 A *Galton board* is a board in which a large number of BB-shots are dropped from a chute at the top of the board and deflected off a number of pins on their way down to the bottom of the board. The final position of each slot is the result of a number of random deflections either to the left or the right. We have written a program **GaltonBoard** to simulate this experiment.

We have run the program for the case of 20 rows of pins and 10,000 shots being dropped. We show the result of this simulation in Figure 3.6.

Note that if we write 0 every time the shot is deflected to the left, and 1 every time it is deflected to the right, then the path of the shot can be described by a sequence of 0's and 1's of length n, just as for the n-fold coin toss.

The distribution shown in Figure 3.6 is an example of an empirical distribution, in the sense that it comes about by means of a sequence of experiments. As expected,

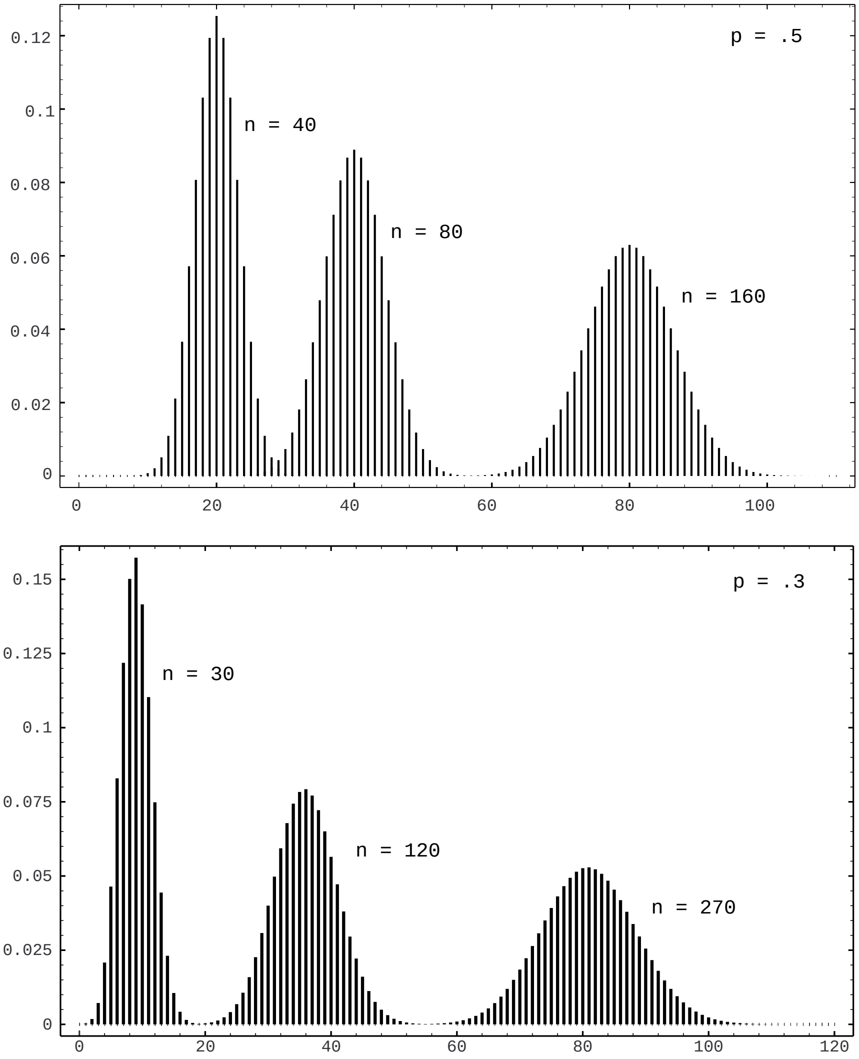

Figure 3.5: Binomial distributions.

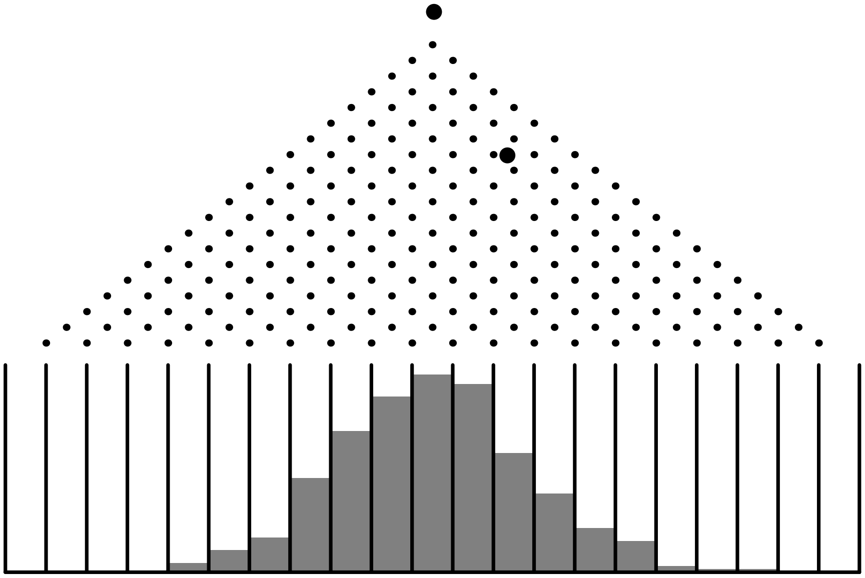

Figure 3.6: Simulation of the Galton board.

this empirical distribution resembles the corresponding binomial distribution with parameters $n = 20$ and $p = 1/2$. □

Hypothesis Testing

Example 3.11 Suppose that ordinary aspirin has been found effective against headaches 60 percent of the time, and that a drug company claims that its new aspirin with a special headache additive is more effective. We can test this claim as follows: we call their claim the *alternate hypothesis,* and its negation, that the additive has no appreciable effect, the *null hypothesis.* Thus the null hypothesis is that $p = .6$, and the alternate hypothesis is that $p > .6$, where p is the probability that the new aspirin is effective.

We give the aspirin to n people to take when they have a headache. We want to find a number m, called the *critical value* for our experiment, such that we reject the null hypothesis if at least m people are cured, and otherwise we accept it. How should we determine this critical value?

First note that we can make two kinds of errors. The first, often called a *type 1 error* in statistics, is to reject the null hypothesis when in fact it is true. The second, called a *type 2 error,* is to accept the null hypothesis when it is false. To determine the probability of both these types of errors we introduce a function $\alpha(p)$, defined to be the probability that we reject the null hypothesis, where this probability is calculated under the assumption that the null hypothesis is true. In the present case, we have

$$\alpha(p) = \sum_{m \le k \le n} b(n, p, k) \ .$$

Note that $\alpha(.6)$ is the probability of a type 1 error, since this is the probability of a high number of successes for an ineffective additive. So for a given n we want to choose m so as to make $\alpha(.6)$ quite small, to reduce the likelihood of a type 1 error. But as m increases above the most probable value $np = .6n$, $\alpha(.6)$, being the upper tail of a binomial distribution, approaches 0. Thus *increasing* m makes a type 1 error less likely.

Now suppose that the additive really is effective, so that p is appreciably greater than .6; say $p = .8$. (This alternative value of p is chosen arbitrarily; the following calculations depend on this choice.) Then choosing m well below $np = .8n$ will increase $\alpha(.8)$, since now $\alpha(.8)$ is all but the lower tail of a binomial distribution. Indeed, if we put $\beta(.8) = 1 - \alpha(.8)$, then $\beta(.8)$ gives us the probability of a type 2 error, and so *decreasing* m makes a type 2 error less likely.

The manufacturer would like to guard against a type 2 error, since if such an error is made, then the test does not show that the new drug is better, when in fact it is. If the alternative value of p is chosen closer to the value of p given in the null hypothesis (in this case $p = .6$), then for a given test population, the value of β will increase. So, if the manufacturer's statistician chooses an alternative value for p which is close to the value in the null hypothesis, then it will be an expensive proposition (i.e., the test population will have to be large) to reject the null hypothesis with a small value of β.

What we hope to do then, for a given test population n, is to choose a value of m, if possible, which makes both these probabilities small. If we make a type 1 error we end up buying a lot of essentially ordinary aspirin at an inflated price; a type 2 error means we miss a bargain on a superior medication. Let us say that we want our critical number m to make each of these undesirable cases less than 5 percent probable.

We write a program **PowerCurve** to plot, for $n = 100$ and selected values of m, the function $\alpha(p)$, for p ranging from .4 to 1. The result is shown in Figure 3.7. We include in our graph a box (in dotted lines) from .6 to .8, with bottom and top at heights .05 and .95. Then a value for m satisfies our requirements if and only if the graph of α enters the box from the bottom, and leaves from the top (why?—which is the type 1 and which is the type 2 criterion?). As m increases, the graph of α moves to the right. A few experiments have shown us that $m = 69$ is the smallest value for m that thwarts a type 1 error, while $m = 73$ is the largest which thwarts a type 2. So we may choose our critical value between 69 and 73. If we're more intent on avoiding a type 1 error we favor 73, and similarly we favor 69 if we regard a type 2 error as worse. Of course, the drug company may not be happy with having as much as a 5 percent chance of an error. They might insist on having a 1 percent chance of an error. For this we would have to increase the number n of trials (see Exercise 28). □

Binomial Expansion

We next remind the reader of an application of the binomial coefficients to algebra. This is the *binomial expansion,* from which we get the term binomial coefficient.

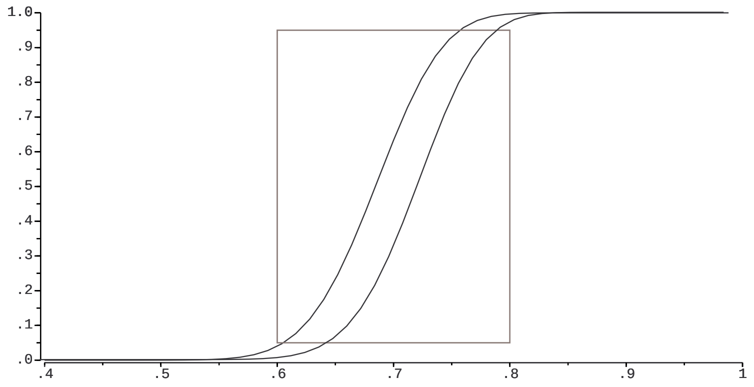

Figure 3.7: The power curve.

Theorem 3.7 (Binomial Theorem) The quantity $(a+b)^n$ can be expressed in the form

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

Proof. To see that this expansion is correct, write

$$(a+b)^n = (a+b)(a+b)\cdots(a+b) \ .$$

When we multiply this out we will have a sum of terms each of which results from a choice of an a or b for each of n factors. When we choose j a's and $(n-j)$ b's, we obtain a term of the form $a^j b^{n-j}$. To determine such a term, we have to specify j of the n terms in the product from which we choose the a. This can be done in $\binom{n}{j}$ ways. Thus, collecting these terms in the sum contributes a term $\binom{n}{j} a^j b^{n-j}$. $\square$

For example, we have

$$\begin{aligned}
(a+b)^0 &= 1 \\
(a+b)^1 &= a+b \\
(a+b)^2 &= a^2+2ab+b^2 \\
(a+b)^3 &= a^3+3a^2b+3ab^2+b^3 \ .
\end{aligned}$$

We see here that the coefficients of successive powers do indeed yield Pascal's triangle.

Corollary 3.1 The sum of the elements in the nth row of Pascal's triangle is 2^n. If the elements in the nth row of Pascal's triangle are added with alternating signs, the sum is 0.

Proof. The first statement in the corollary follows from the fact that

$$2^n = (1+1)^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} ,$$

and the second from the fact that

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

□

The first statement of the corollary tells us that the number of subsets of a set of n elements is 2^n. We shall use the second statement in our next application of the binomial theorem.

We have seen that, when A and B are any two events (cf. Section 1.2),

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

We now extend this theorem to a more general version, which will enable us to find the probability that at least one of a number of events occurs.

Inclusion-Exclusion Principle

Theorem 3.8 Let P be a probability measure on a sample space Ω, and let $\{A_1,\ A_2,\ \ldots,\ A_n\}$ be a finite set of events. Then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^n P(A_i) \ - \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$

$$+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \cdots . \quad (3.3)$$

That is, to find the probability that at least one of n events A_i occurs, first add the probability of each event, then subtract the probabilities of all possible two-way intersections, add the probability of all three-way intersections, and so forth.

Proof. If the outcome ω occurs in at least one of the events A_i, its probability is added exactly once by the left side of Equation 3.3. We must show that it is added exactly once by the right side of Equation 3.3. Assume that ω is in exactly k of the sets. Then its probability is added k times in the first term, subtracted $\binom{k}{2}$ times in the second, added $\binom{k}{3}$ times in the third term, and so forth. Thus, the total number of times that it is added is

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \cdots (-1)^{k-1} \binom{k}{k} .$$

But

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

Hence,

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

If the outcome ω is not in any of the events A_i, then it is not counted on either side of the equation. □

Hat Check Problem

Example 3.12 We return to the hat check problem discussed in Section 3.1, that is, the problem of finding the probability that a random permutation contains at least one fixed point. Recall that a permutation is a one-to-one map of a set $A = \{a_1, a_2, \ldots, a_n\}$ onto itself. Let A_i be the event that the ith element a_i remains fixed under this map. If we require that a_i is fixed, then the map of the remaining $n-1$ elements provides an arbitrary permutation of $(n-1)$ objects. Since there are $(n-1)!$ such permutations, $P(A_i) = (n-1)!/n! = 1/n$. Since there are n choices for a_i, the first term of Equation 3.3 is 1. In the same way, to have a particular pair (a_i, a_j) fixed, we can choose any permutation of the remaining $n-2$ elements; there are $(n-2)!$ such choices and thus

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \ .$$

The number of terms of this form in the right side of Equation 3.3 is

$$\binom{n}{2} = \frac{n(n-1)}{2!} \ .$$

Hence, the second term of Equation 3.3 is

$$-\frac{n(n-1)}{2!} \cdot \frac{1}{n(n-1)} = -\frac{1}{2!} \ .$$

Similarly, for any specific three events A_i, A_j, A_k,

$$P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)} \ ,$$

and the number of such terms is

$$\binom{n}{3} = \frac{n(n-1)(n-2)}{3!} \ ,$$

making the third term of Equation 3.3 equal to 1/3!. Continuing in this way, we obtain

$$P(\text{at least one fixed point}) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots (-1)^{n-1} \frac{1}{n!}$$

and

$$P(\text{no fixed point}) = \frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^n \frac{1}{n!} \ .$$

n	Probability that no one gets his own hat back
3	.333333
4	.375
5	.366667
6	.368056
7	.367857
8	.367882
9	.367879
10	.367879

Table 3.9: Hat check problem.

From calculus we learn that

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n + \cdots \ .$$

Thus, if $x = -1$, we have

$$\begin{aligned} e^{-1} &= \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} + \cdots \\ &= .3678794\ . \end{aligned}$$

Therefore, the probability that there is no fixed point, i.e., that none of the n people gets his own hat back, is equal to the sum of the first n terms in the expression for e^{-1}. This series converges very fast. Calculating the partial sums for $n = 3$ to 10 gives the data in Table 3.9.

After $n = 9$ the probabilities are essentially the same to six significant figures. Interestingly, the probability of no fixed point alternately increases and decreases as n increases. Finally, we note that our exact results are in good agreement with our simulations reported in the previous section. □

Choosing a Sample Space

We now have some of the tools needed to accurately describe sample spaces and to assign probability functions to those sample spaces. Nevertheless, in some cases, the description and assignment process is somewhat arbitrary. Of course, it is to be hoped that the description of the sample space and the subsequent assignment of a probability function will yield a model which accurately predicts what would happen if the experiment were actually carried out. As the following examples show, there are situations in which "reasonable" descriptions of the sample space do not produce a model which fits the data.

In Feller's book,[14] a pair of models is given which describe arrangements of certain kinds of elementary particles, such as photons and protons. It turns out that experiments have shown that certain types of elementary particles exhibit behavior

[14] W. Feller, *Introduction to Probability Theory and Its Applications* vol. 1, 3rd ed. (New York: John Wiley and Sons, 1968), p. 41

which is accurately described by one model, called *"Bose-Einstein statistics,"* while other types of elementary particles can be modelled using *"Fermi-Dirac statistics."* Feller says:

> We have here an instructive example of the impossibility of selecting or justifying probability models by *a priori* arguments. In fact, no pure reasoning could tell that photons and protons would not obey the same probability laws.

We now give some examples of this description and assignment process.

Example 3.13 In the quantum mechanical model of the helium atom, various parameters can be used to classify the energy states of the atom. In the triplet spin state ($S = 1$) with orbital angular momentum 1 ($L = 1$), there are three possibilities, 0, 1, or 2, for the total angular momentum (J). (It is not assumed that the reader knows what any of this means; in fact, the example is more illustrative if the reader does *not* know anything about quantum mechanics.) We would like to assign probabilities to the three possibilities for J. The reader is undoubtedly resisting the idea of assigning the probability of $1/3$ to each of these outcomes. She should now ask herself why she is resisting this assignment. The answer is probably because she does not have any "intuition" (i.e., experience) about the way in which helium atoms behave. In fact, in this example, the probabilities $1/9$, $3/9$, and $5/9$ are assigned by the theory. The theory gives these assignments because these frequencies were observed *in experiments* and further parameters were developed in the theory to allow these frequencies to be predicted. □

Example 3.14 Suppose two pennies are flipped once each. There are several "reasonable" ways to describe the sample space. One way is to count the number of heads in the outcome; in this case, the sample space can be written $\{0, 1, 2\}$. Another description of the sample space is the set of all ordered pairs of H's and T's, i.e.,

$$\{(H, H), (H, T), (T, H), (T, T)\}.$$

Both of these descriptions are accurate ones, but it is easy to see that (at most) one of these, if assigned a constant probability function, can claim to accurately model reality. In this case, as opposed to the preceding example, the reader will probably say that the second description, with each outcome being assigned a probability of $1/4$, is the "right" description. This conviction is due to experience; there is no proof that this is the way reality works. □

The reader is also referred to Exercise 26 for another example of this process.

Historical Remarks

The binomial coefficients have a long and colorful history leading up to Pascal's *Treatise on the Arithmetical Triangle,*[15] where Pascal developed many important

[15] B. Pascal, *Traité du Triangle Arithmétique* (Paris: Desprez, 1665).

1	1	1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	
1	3	6	10	15	21	28	36		
1	4	10	20	35	56	84			
1	5	15	35	70	126				
1	6	21	56	126					
1	7	28	84						
1	8	36							
1	9								
1									

Table 3.10: Pascal's triangle.

natural numbers	1	2	3	4	5	6	7	8	9
triangular numbers	1	3	6	10	15	21	28	36	45
tetrahedral numbers	1	4	10	20	35	56	84	120	165

Table 3.11: Figurate numbers.

properties of these numbers. This history is set forth in the book *Pascal's Arithmetical Triangle* by A. W. F. Edwards.[16] Pascal wrote his triangle in the form shown in Table 3.10.

Edwards traces three different ways that the binomial coefficients arose. He refers to these as the *figurate numbers,* the *combinatorial numbers,* and the *binomial numbers.* They are all names for the same thing (which we have called binomial coefficients) but that they are all the same was not appreciated until the sixteenth century.

The *figurate numbers* date back to the Pythagorean interest in number patterns around 540 BC. The Pythagoreans considered, for example, triangular patterns shown in Figure 3.8. The sequence of numbers

$$1, 3, 6, 10, \ldots$$

obtained as the number of points in each triangle are called *triangular numbers.* From the triangles it is clear that the nth triangular number is simply the sum of the first n integers. The *tetrahedral numbers* are the sums of the triangular numbers and were obtained by the Greek mathematicians Theon and Nicomachus at the beginning of the second century BC. The tetrahedral number 10, for example, has the geometric representation shown in Figure 3.9. The first three types of figurate numbers can be represented in tabular form as shown in Table 3.11.

These numbers provide the first four rows of Pascal's triangle, but the table was not to be completed in the West until the sixteenth century.

In the East, Hindu mathematicians began to encounter the binomial coefficients in combinatorial problems. Bhaskara in his *Lilavati* of 1150 gave a rule to find the

[16] A. W. F. Edwards, *Pascal's Arithmetical Triangle* (London: Griffin, 1987).

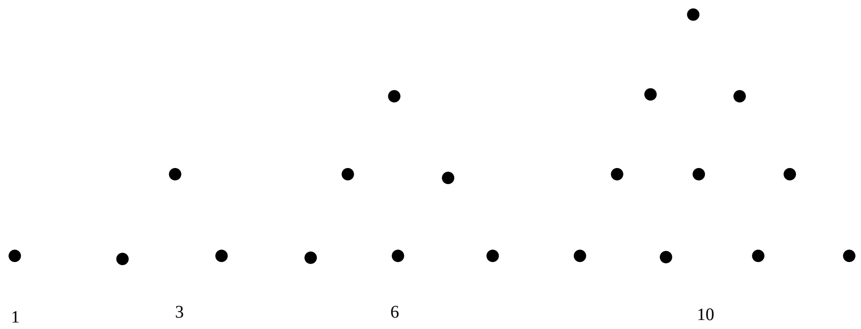

Figure 3.8: Pythagorean triangular patterns.

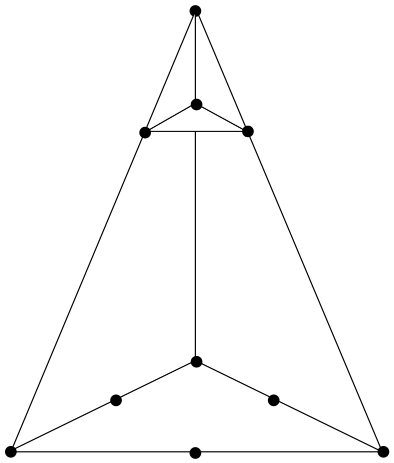

Figure 3.9: Geometric representation of the tetrahedral number 10.

11					
12	22				
13	23	33			
14	24	34	44		
15	25	35	45	55	
16	26	36	46	56	66

Table 3.12: Outcomes for the roll of two dice.

number of medicinal preparations using 1, 2, 3, 4, 5, or 6 possible ingredients.[17] His rule is equivalent to our formula

$$\binom{n}{r} = \frac{(n)_r}{r!} \ .$$

The binomial numbers as coefficients of $(a+b)^n$ appeared in the works of mathematicians in China around 1100. There are references about this time to "the tabulation system for unlocking binomial coefficients." The triangle to provide the coefficients up to the eighth power is given by Chu Shih-chieh in a book written around 1303 (see Figure 3.10).[18] The original manuscript of Chu's book has been lost, but copies have survived. Edwards notes that there is an error in this copy of Chu's triangle. Can you find it? (*Hint*: Two numbers which should be equal are not.) Other copies do not show this error.

The first appearance of Pascal's triangle in the West seems to have come from calculations of Tartaglia in calculating the number of possible ways that n dice might turn up.[19] For one die the answer is clearly 6. For two dice the possibilities may be displayed as shown in Table 3.12.

Displaying them this way suggests the sixth triangular number $1+2+3+4+5+6=21$ for the throw of 2 dice. Tartaglia "on the first day of Lent, 1523, in Verona, having thought about the problem all night,"[20] realized that the extension of the figurate table gave the answers for n dice. The problem had suggested itself to Tartaglia from watching people casting their own horoscopes by means of a *Book of Fortune,* selecting verses by a process which included noting the numbers on the faces of three dice. The 56 ways that three dice can fall were set out on each page. The way the numbers were written in the book did not suggest the connection with figurate numbers, but a method of enumeration similar to the one we used for 2 dice does. Tartaglia's table was not published until 1556.

A table for the binomial coefficients was published in 1554 by the German mathematician Stifel.[21] Pascal's triangle appears also in Cardano's *Opus novum* of 1570.[22]

[17] ibid., p. 27.

[18] J. Needham, *Science and Civilization in China,* vol. 3 (New York: Cambridge University Press, 1959), p. 135.

[19] N. Tartaglia, *General Trattato di Numeri et Misure* (Vinegia, 1556).

[20] Quoted in Edwards, op. cit., p. 37.

[21] M. Stifel, *Arithmetica Integra* (Norimburgae, 1544).

[22] G. Cardano, *Opus Novum de Proportionibus Numerorum* (Basilea, 1570).

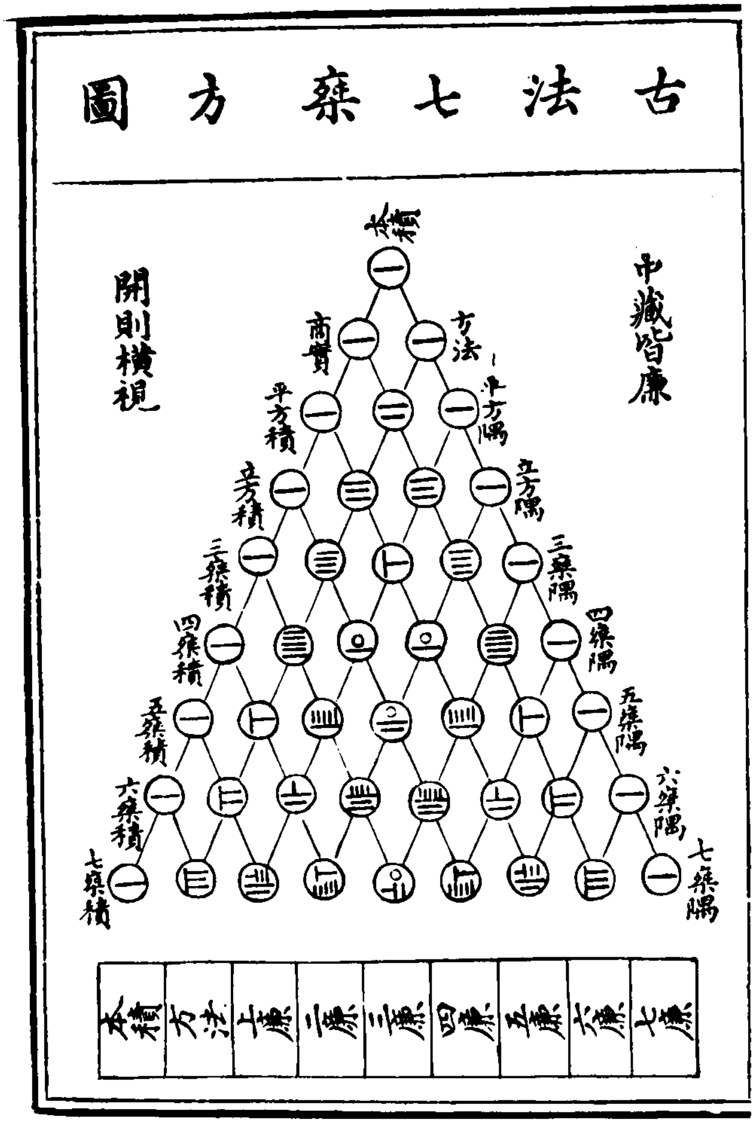

Figure 3.10: Chu Shih-chieh's triangle. [From J. Needham, *Science and Civilization in China,* vol. 3 (New York: Cambridge University Press, 1959), p. 135. Reprinted with permission.]

Cardano was interested in the problem of finding the number of ways to choose r objects out of n. Thus by the time of Pascal's work, his triangle had appeared as a result of looking at the figurate numbers, the combinatorial numbers, and the binomial numbers, and the fact that all three were the same was presumably pretty well understood.

Pascal's interest in the binomial numbers came from his letters with Fermat concerning a problem known as the problem of points. This problem, and the correspondence between Pascal and Fermat, were discussed in Chapter 1. The reader will recall that this problem can be described as follows: Two players A and B are playing a sequence of games and the first player to win n games wins the match. It is desired to find the probability that A wins the match at a time when A has won a games and B has won b games. (See Exercises 4.1.40-4.1.42.)

Pascal solved the problem by backward induction, much the way we would do today in writing a computer program for its solution. He referred to the combinatorial method of Fermat which proceeds as follows: If A needs c games and B needs d games to win, we require that the players continue to play until they have played $c + d - 1$ games. The winner in this extended series will be the same as the winner in the original series. The probability that A wins in the extended series and hence in the original series is

$$\sum_{r=c}^{c+d-1} \frac{1}{2^{c+d-1}} \binom{c+d-1}{r} .$$

Even at the time of the letters Pascal seemed to understand this formula.

Suppose that the first player to win n games wins the match, and suppose that each player has put up a stake of x. Pascal studied the value of winning a particular game. By this he meant the increase in the expected winnings of the winner of the particular game under consideration. He showed that the value of the first game is

$$\frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)}{2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2n)} x .$$

His proof of this seems to use Fermat's formula and the fact that the above ratio of products of odd to products of even numbers is equal to the probability of exactly n heads in $2n$ tosses of a coin. (See Exercise 39.)

Pascal presented Fermat with the table shown in Table 3.13. He states:

> You will see as always, that the value of the first game is equal to that of the second which is easily shown by combinations. You will see, in the same way, that the numbers in the first line are always increasing; so also are those in the second; and those in the third. But those in the fourth line are decreasing, and those in the fifth, etc. This seems odd.[23]

The student can pursue this question further using the computer and Pascal's backward iteration method for computing the expected payoff at any point in the series.

[23] F. N. David, op. cit., p. 235.

	if each one staken 256 in					
From my opponent's 256 positions I get, for the	6 games	5 games	4 games	3 games	2 games	1 games
1st game	63	70	80	96	128	256
2nd game	63	70	80	96	128	
3rd game	56	60	64	64		
4th game	42	40	32			
5th game	24	16				
6th game	8					

Table 3.13: Pascal's solution for the problem of points.

In his treatise, Pascal gave a formal proof of Fermat's combinatorial formula as well as proofs of many other basic properties of binomial numbers. Many of his proofs involved induction and represent some of the first proofs by this method. His book brought together all the different aspects of the numbers in the Pascal triangle as known in 1654, and, as Edwards states, "That the Arithmetical Triangle should bear Pascal's name cannot be disputed."[24]

The first serious study of the binomial distribution was undertaken by James Bernoulli in his *Ars Conjectandi* published in 1713.[25] We shall return to this work in the historical remarks in Chapter 8.

Exercises

1 Compute the following:

(a) $\binom{6}{3}$

(b) $b(5, .2, 4)$

(c) $\binom{7}{2}$

(d) $\binom{26}{26}$

(e) $b(4, .2, 3)$

(f) $\binom{6}{2}$

(g) $\binom{10}{9}$

(h) $b(8, .3, 5)$

2 In how many ways can we choose five people from a group of ten to form a committee?

3 How many seven-element subsets are there in a set of nine elements?

4 Using the relation Equation 3.1 write a program to compute Pascal's triangle, putting the results in a matrix. Have your program print the triangle for $n = 10$.

[24] A. W. F. Edwards, op. cit., p. ix.

[25] J. Bernoulli, *Ars Conjectandi* (Basil: Thurnisiorum, 1713).

5 Use the program **BinomialProbabilities** to find the probability that, in 100 tosses of a fair coin, the number of heads that turns up lies between 35 and 65, between 40 and 60, and between 45 and 55.

6 Charles claims that he can distinguish between beer and ale 75 percent of the time. Ruth bets that he cannot and, in fact, just guesses. To settle this, a bet is made: Charles is to be given ten small glasses, each having been filled with beer or ale, chosen by tossing a fair coin. He wins the bet if he gets seven or more correct. Find the probability that Charles wins if he has the ability that he claims. Find the probability that Ruth wins if Charles is guessing.

7 Show that

$$b(n,p,j) = \frac{p}{q}\left(\frac{n-j+1}{j}\right) b(n,p,j-1)\ ,$$

for $j \geq 1$. Use this fact to determine the value or values of j which give $b(n,p,j)$ its greatest value. *Hint*: Consider the successive ratios as j increases.

8 A die is rolled 30 times. What is the probability that a 6 turns up exactly 5 times? What is the most probable number of times that a 6 will turn up?

9 Find integers n and r such that the following equation is true:

$$\binom{13}{5} + 2\binom{13}{6} + \binom{13}{7} = \binom{n}{r}\ .$$

10 In a ten-question true-false exam, find the probability that a student gets a grade of 70 percent or better by guessing. Answer the same question if the test has 30 questions, and if the test has 50 questions.

11 A restaurant offers apple and blueberry pies and stocks an equal number of each kind of pie. Each day ten customers request pie. They choose, with equal probabilities, one of the two kinds of pie. How many pieces of each kind of pie should the owner provide so that the probability is about .95 that each customer gets the pie of his or her own choice?

12 A poker hand is a set of 5 cards randomly chosen from a deck of 52 cards. Find the probability of a

(a) royal flush (ten, jack, queen, king, ace in a single suit).

(b) straight flush (five in a sequence in a single suit, but not a royal flush).

(c) four of a kind (four cards of the same face value).

(d) full house (one pair and one triple, each of the same face value).

(e) flush (five cards in a single suit but not a straight or royal flush).

(f) straight (five cards in a sequence, not all the same suit). (Note that in straights, an ace counts high or low.)

13 If a set has $2n$ elements, show that it has more subsets with n elements than with any other number of elements.

14 Let $b(2n, .5, n)$ be the probability that in $2n$ tosses of a fair coin exactly n heads turn up. Using Stirling's formula (Theorem 3.3), show that $b(2n, .5, n) \sim 1/\sqrt{\pi n}$. Use the program **BinomialProbabilities** to compare this with the exact value for $n = 10$ to 25.

15 A baseball player, Smith, has a batting average of .300 and in a typical game comes to bat three times. Assume that Smith's hits in a game can be considered to be a Bernoulli trials process with probability .3 for *success.* Find the probability that Smith gets 0, 1, 2, and 3 hits.

16 The Siwash University football team plays eight games in a season, winning three, losing three, and ending two in a tie. Show that the number of ways that this can happen is

$$\binom{8}{3}\binom{5}{3} = \frac{8!}{3!\,3!\,2!}\ .$$

17 Using the technique of Exercise 16, show that the number of ways that one can put n different objects into three boxes with a in the first, b in the second, and c in the third is $n!/(a!\,b!\,c!)$.

18 Baumgartner, Prosser, and Crowell are grading a calculus exam. There is a true-false question with ten parts. Baumgartner notices that one student has only two out of the ten correct and remarks, "The student was not even bright enough to have flipped a coin to determine his answers." "Not so clear," says Prosser. "With 340 students I bet that if they all flipped coins to determine their answers there would be at least one exam with two or fewer answers correct." Crowell says, "I'm with Prosser. In fact, I bet that we should expect at least one exam in which no answer is correct if everyone is just guessing." Who is right in all of this?

19 A gin hand consists of 10 cards from a deck of 52 cards. Find the probability that a gin hand has

(a) all 10 cards of the same suit.

(b) exactly 4 cards in one suit and 3 in two other suits.

(c) a 4, 3, 2, 1, distribution of suits.

20 A six-card hand is dealt from an ordinary deck of cards. Find the probability that:

(a) All six cards are hearts.

(b) There are three aces, two kings, and one queen.

(c) There are three cards of one suit and three of another suit.

21 A lady wishes to color her fingernails on one hand using at most two of the colors red, yellow, and blue. How many ways can she do this?

22 How many ways can six indistinguishable letters be put in three mail boxes? *Hint*: One representation of this is given by a sequence |LL|L|LLL| where the |'s represent the partitions for the boxes and the L's the letters. Any possible way can be so described. Note that we need two bars at the ends and the remaining two bars and the six L's can be put in any order.

23 Using the method for the hint in Exercise 22, show that r indistinguishable objects can be put in n boxes in

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

different ways.

24 A travel bureau estimates that when 20 tourists go to a resort with ten hotels they distribute themselves as if the bureau were putting 20 indistinguishable objects into ten distinguishable boxes. Assuming this model is correct, find the probability that no hotel is left vacant when the first group of 20 tourists arrives.

25 An elevator takes on six passengers and stops at ten floors. We can assign two different equiprobable measures for the ways that the passengers are discharged: (a) we consider the passengers to be distinguishable or (b) we consider them to be indistinguishable (see Exercise 23 for this case). For each case, calculate the probability that all the passengers get off at different floors.

26 You are playing *heads or tails* with Prosser but you suspect that his coin is unfair. Von Neumann suggested that you proceed as follows: Toss Prosser's coin twice. If the outcome is HT call the result *win.* if it is TH call the result *lose.* If it is TT or HH ignore the outcome and toss Prosser's coin twice again. Keep going until you get either an HT or a TH and call the result win or lose in a single play. Repeat this procedure for each play. Assume that Prosser's coin turns up heads with probability p.

(a) Find the probability of HT, TH, HH, TT with two tosses of Prosser's coin.

(b) Using part (a), show that the probability of a win on any one play is 1/2, no matter what p is.

27 John claims that he has extrasensory powers and can tell which of two symbols is on a card turned face down (see Example 3.11). To test his ability he is asked to do this for a sequence of trials. Let the null hypothesis be that he is just guessing, so that the probability is 1/2 of his getting it right each time, and let the alternative hypothesis be that he can name the symbol correctly more than half the time. Devise a test with the property that the probability of a type 1 error is less than .05 and the probability of a type 2 error is less than .05 if John can name the symbol correctly 75 percent of the time.

28 In Example 3.11 assume the alternative hypothesis is that $p = .8$ and that it is desired to have the probability of each type of error less than .01. Use the program **PowerCurve** to determine values of n and m that will achieve this. Choose n as small as possible.

29 A drug is assumed to be effective with an unknown probability p. To estimate p the drug is given to n patients. It is found to be effective for m patients. The *method of maximum likelihood* for estimating p states that we should choose the value for p that gives the highest probability of getting what we got on the experiment. Assuming that the experiment can be considered as a Bernoulli trials process with probability p for success, show that the maximum likelihood estimate for p is the proportion m/n of successes.

30 Recall that in the World Series the first team to win four games wins the series. The series can go at most seven games. Assume that the Red Sox and the Mets are playing the series. Assume that the Mets win each game with probability p. Fermat observed that even though the series might not go seven games, the probability that the Mets win the series is the same as the probability that they win four or more game in a series that was forced to go seven games no matter who wins the individual games.

(a) Using the program **PowerCurve** of Example 3.11 find the probability that the Mets win the series for the cases $p = .5$, $p = .6$, $p = .7$.

(b) Assume that the Mets have probability .6 of winning each game. Use the program **PowerCurve** to find a value of n so that, if the series goes to the first team to win more than half the games, the Mets will have a 95 percent chance of winning the series. Choose n as small as possible.

31 Each of the four engines on an airplane functions correctly on a given flight with probability .99, and the engines function independently of each other. Assume that the plane can make a safe landing if at least two of its engines are functioning correctly. What is the probability that the engines will allow for a safe landing?

32 A small boy is lost coming down Mount Washington. The leader of the search team estimates that there is a probability p that he came down on the east side and a probability $1 - p$ that he came down on the west side. He has n people in his search team who will search independently and, if the boy is on the side being searched, each member will find the boy with probability u. Determine how he should divide the n people into two groups to search the two sides of the mountain so that he will have the highest probability of finding the boy. How does this depend on u?

***33** $2n$ balls are chosen at random from a total of $2n$ red balls and $2n$ blue balls. Find a combinatorial expression for the probability that the chosen balls are equally divided in color. Use Stirling's formula to estimate this probability.

Using **BinomialProbabilities**, compare the exact value with Stirling's approximation for $n = 20$.

34 Assume that every time you buy a box of Wheaties, you receive one of the pictures of the n players on the New York Yankees. Over a period of time, you buy $m \geq n$ boxes of Wheaties.

(a) Use Theorem 3.8 to show that the probability that you get all n pictures is

$$\begin{aligned} 1 \quad &- \quad \binom{n}{1}\left(\frac{n-1}{n}\right)^m + \binom{n}{2}\left(\frac{n-2}{n}\right)^m - \cdots \\ &+ \quad (-1)^{n-1}\binom{n}{n-1}\left(\frac{1}{n}\right)^m . \end{aligned}$$

Hint: Let E_k be the event that you do not get the kth player's picture.

(b) Write a computer program to compute this probability. Use this program to find, for given n, the smallest value of m which will give probability $\geq .5$ of getting all n pictures. Consider $n = 50$, 100, and 150 and show that $m = n \log n + n \log 2$ is a good estimate for the number of boxes needed. (For a derivation of this estimate, see Feller.[26])

***35** Prove the following *binomial identity*

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

Hint: Consider an urn with n red balls and n blue balls inside. Show that each side of the equation equals the number of ways to choose n balls from the urn.

36 Let j and n be positive integers, with $j \leq n$. An experiment consists of choosing, at random, a j-tuple of *positive* integers whose sum is at most n.

(a) Find the size of the sample space. *Hint*: Consider n indistinguishable balls placed in a row. Place j markers between consecutive pairs of balls, with no two markers between the same pair of balls. (We also allow one of the n markers to be placed at the end of the row of balls.) Show that there is a 1-1 correspondence between the set of possible positions for the markers and the set of j-tuples whose size we are trying to count.

(b) Find the probability that the j-tuple selected contains at least one 1.

37 Let n (mod m) denote the remainder when the integer n is divided by the integer m. Write a computer program to compute the numbers $\binom{n}{j}$ (mod m) where $\binom{n}{j}$ is a binomial coefficient and m is an integer. You can do this by using the recursion relations for generating binomial coefficients, doing all the

[26] W. Feller, *Introduction to Probability Theory and its Applications,* vol. I, 3rd ed. (New York: John Wiley & Sons, 1968), p. 106.

arithmetic using the basic function mod(n, m). Try to write your program to make as large a table as possible. Run your program for the cases $m = 2$ to 7. Do you see any patterns? In particular, for the case $m = 2$ and n a power of 2, verify that all the entries in the $(n-1)$st row are 1. (The corresponding binomial numbers are odd.) Use your pictures to explain why this is true.

38 Lucas[27] proved the following general result relating to Exercise 37. If p is any prime number, then $\binom{n}{j}$ (mod p) can be found as follows: Expand n and j in base p as $n = s_0 + s_1 p + s_2 p^2 + \cdots + s_k p^k$ and $j = r_0 + r_1 p + r_2 p^2 + \cdots + r_k p^k$, respectively. (Here k is chosen large enough to represent all numbers from 0 to n in base p using k digits.) Let $s = (s_0, s_1, s_2, \ldots, s_k)$ and $r = (r_0, r_1, r_2, \ldots, r_k)$. Then

$$\binom{n}{j} \pmod{p} = \prod_{i=0}^{k} \binom{s_i}{r_i} \pmod{p} \ .$$

For example, if $p = 7$, $n = 12$, and $j = 9$, then

$$\begin{aligned} 12 &= 5 \cdot 7^0 + 1 \cdot 7^1 \ , \\ 9 &= 2 \cdot 7^0 + 1 \cdot 7^1 \ , \end{aligned}$$

so that

$$\begin{aligned} s &= (5, 1) \ , \\ r &= (2, 1) \ , \end{aligned}$$

and this result states that

$$\binom{12}{9} \pmod{p} = \binom{5}{2}\binom{1}{1} \pmod{7} \ .$$

Since $\binom{12}{9} = 220 = 3 \pmod{7}$, and $\binom{5}{2} = 10 = 3 \pmod{7}$, we see that the result is correct for this example.

Show that this result implies that, for $p = 2$, the $(p^k - 1)$st row of your triangle in Exercise 37 has no zeros.

39 Prove that the probability of exactly n heads in $2n$ tosses of a fair coin is given by the product of the odd numbers up to $2n - 1$ divided by the product of the even numbers up to $2n$.

40 Let n be a positive integer, and assume that j is a positive integer not exceeding $n/2$. Show that in Theorem 3.5, if one alternates the multiplications and divisions, then all of the intermediate values in the calculation are integers. Show also that none of these intermediate values exceed the final value.

[27] E. Lucas, "Théorie des Functions Numériques Simplement Periodiques," *American J. Math.*, vol. 1 (1878), pp. 184-240, 289-321.

3.3 Card Shuffling

Much of this section is based upon an article by Brad Mann,[28] which is an exposition of an article by David Bayer and Persi Diaconis.[29]

Riffle Shuffles

Given a deck of n cards, how many times must we shuffle it to make it "random"? Of course, the answer depends upon the method of shuffling which is used and what we mean by "random." We shall begin the study of this question by considering a standard model for the riffle shuffle.

We begin with a deck of n cards, which we will assume are labelled in increasing order with the integers from 1 to n. A riffle shuffle consists of a cut of the deck into two stacks and an interleaving of the two stacks. For example, if $n = 6$, the initial ordering is $(1, 2, 3, 4, 5, 6)$, and a cut might occur between cards 2 and 3. This gives rise to two stacks, namely $(1, 2)$ and $(3, 4, 5, 6)$. These are interleaved to form a new ordering of the deck. For example, these two stacks might form the ordering $(1, 3, 4, 2, 5, 6)$. In order to discuss such shuffles, we need to assign a probability measure to the set of all possible shuffles. There are several reasonable ways in which this can be done. We will give several different assignment strategies, and show that they are equivalent. (This does not mean that this assignment is the only reasonable one.) First, we assign the binomial probability $b(n, 1/2, k)$ to the event that the cut occurs after the kth card. Next, we assume that all possible interleavings, given a cut, are equally likely. Thus, to complete the assignment of probabilities, we need to determine the number of possible interleavings of two stacks of cards, with k and $n-k$ cards, respectively.

We begin by writing the second stack in a line, with spaces in between each pair of consecutive cards, and with spaces at the beginning and end (so there are $n - k + 1$ spaces). We choose, with replacement, k of these spaces, and place the cards from the first stack in the chosen spaces. This can be done in

$$\binom{n}{k}$$

ways. Thus, the probability of a given interleaving should be

$$\frac{1}{\binom{n}{k}} .$$

Next, we note that if the new ordering is not the identity ordering, it is the result of a unique cut-interleaving pair. If the new ordering is the identity, it is the result of any one of $n + 1$ cut-interleaving pairs.

We define a *rising sequence* in an ordering to be a maximal subsequence of consecutive integers in increasing order. For example, in the ordering

$$(2, 3, 5, 1, 4, 7, 6) ,$$

[28] B. Mann, "How Many Times Should You Shuffle a Deck of Cards?", *UMAP Journal*, vol. 15, no. 4 (1994), pp. 303–331.

[29] D. Bayer and P. Diaconis, "Trailing the Dovetail Shuffle to its Lair," *Annals of Applied Probability*, vol. 2, no. 2 (1992), pp. 294–313.

there are 4 rising sequences; they are (1), $(2, 3, 4)$, $(5, 6)$, and (7). It is easy to see that an ordering is the result of a riffle shuffle applied to the identity ordering if and only if it has no more than two rising sequences. (If the ordering has two rising sequences, then these rising sequences correspond to the two stacks induced by the cut, and if the ordering has one rising sequence, then it is the identity ordering.) Thus, the sample space of orderings obtained by applying a riffle shuffle to the identity ordering is naturally described as the set of all orderings with at most two rising sequences.

It is now easy to assign a probability measure to this sample space. Each ordering with two rising sequences is assigned the value

$$\frac{b(n, 1/2, k)}{\binom{n}{k}} = \frac{1}{2^n} \ ,$$

and the identity ordering is assigned the value

$$\frac{n+1}{2^n} \ .$$

There is another way to view a riffle shuffle. We can imagine starting with a deck cut into two stacks as before, with the same probabilities assignment as before i.e., the binomial distribution. Once we have the two stacks, we take cards, one by one, off of the bottom of the two stacks, and place them onto one stack. If there are k_1 and k_2 cards, respectively, in the two stacks at some point in this process, then we make the assumption that the probabilities that the next card to be taken comes from a given stack is proportional to the current stack size. This implies that the probability that we take the next card from the first stack equals

$$\frac{k_1}{k_1 + k_2} \ ,$$

and the corresponding probability for the second stack is

$$\frac{k_2}{k_1 + k_2} \ .$$

We shall now show that this process assigns the uniform probability to each of the possible interleavings of the two stacks.

Suppose, for example, that an interleaving came about as the result of choosing cards from the two stacks in some order. The probability that this result occurred is the product of the probabilities at each point in the process, since the choice of card at each point is assumed to be independent of the previous choices. Each factor of this product is of the form

$$\frac{k_i}{k_1 + k_2} \ ,$$

where $i = 1$ or 2, and the denominator of each factor equals the number of cards left to be chosen. Thus, the denominator of the probability is just $n!$. At the moment when a card is chosen from a stack that has i cards in it, the numerator of the corresponding factor in the probability is i, and the number of cards in this stack decreases by 1. Thus, the numerator is seen to be $k!(n-k)!$, since all cards in both stacks are eventually chosen. Therefore, this process assigns the probability

$$\frac{1}{\binom{n}{k}}$$

to each possible interleaving.

We now turn to the question of what happens when we riffle shuffle s times. It should be clear that if we start with the identity ordering, we obtain an ordering with at most 2^s rising sequences, since a riffle shuffle creates at most two rising sequences from every rising sequence in the starting ordering. In fact, it is not hard to see that each such ordering is the result of s riffle shuffles. The question becomes, then, in how many ways can an ordering with r rising sequences come about by applying s riffle shuffles to the identity ordering? In order to answer this question, we turn to the idea of an a-shuffle.

a-Shuffles

There are several ways to visualize an a-shuffle. One way is to imagine a creature with a hands who is given a deck of cards to riffle shuffle. The creature naturally cuts the deck into a stacks, and then riffles them together. (Imagine that!) Thus, the ordinary riffle shuffle is a 2-shuffle. As in the case of the ordinary 2-shuffle, we allow some of the stacks to have 0 cards. Another way to visualize an a-shuffle is to think about its inverse, called an a-unshuffle. This idea is described in the proof of the next theorem.

We will now show that an a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle. This means, in particular, that s riffle shuffles in succession are equivalent to one 2^s-shuffle. This equivalence is made precise by the following theorem.

Theorem 3.9 Let a and b be two positive integers. Let $S_{a,b}$ be the set of all ordered pairs in which the first entry is an a-shuffle and the second entry is a b-shuffle. Let S_{ab} be the set of all ab-shuffles. Then there is a 1-1 correspondence between $S_{a,b}$ and S_{ab} with the following property. Suppose that (T_1, T_2) corresponds to T_3. If T_1 is applied to the identity ordering, and T_2 is applied to the resulting ordering, then the final ordering is the same as the ordering that is obtained by applying T_3 to the identity ordering.

Proof. The easiest way to describe the required correspondence is through the idea of an unshuffle. An a-unshuffle begins with a deck of n cards. One by one, cards are taken from the top of the deck and placed, with equal probability, on the bottom of any one of a stacks, where the stacks are labelled from 0 to $a-1$. After all of the cards have been distributed, we combine the stacks to form one stack by placing stack i on top of stack $i+1$, for $0 \leq i \leq a-1$. It is easy to see that if one starts with a deck, there is exactly one way to cut the deck to obtain the a stacks generated by the a-unshuffle, and with these a stacks, there is exactly one way to interleave them to obtain the deck in the order that it was in before the unshuffle was performed. Thus, this a-unshuffle corresponds to a unique a-shuffle, and this a-shuffle is the inverse of the original a-unshuffle.

If we apply an ab-unshuffle U_3 to a deck, we obtain a set of ab stacks, which are then combined, in order, to form one stack. We label these stacks with ordered pairs of integers, where the first coordinate is between 0 and $a-1$, and the second coordinate is between 0 and $b-1$. Then we label each card with the label of its stack. The number of possible labels is ab, as required. Using this labelling, we can describe how to find a b-unshuffle and an a-unshuffle, such that if these two unshuffles are applied in this order to the deck, we obtain the same set of ab stacks as were obtained by the ab-unshuffle.

To obtain the b-unshuffle U_2, we sort the deck into b stacks, with the ith stack containing all of the cards with second coordinate i, for $0 \leq i \leq b-1$. Then these stacks are combined to form one stack. The a-unshuffle U_1 proceeds in the same manner, except that the first coordinates of the labels are used. The resulting a stacks are then combined to form one stack.

The above description shows that the cards ending up on top are all those labelled $(0,0)$. These are followed by those labelled $(0,1)$, $(0,2)$, $\ldots$, $(0,b-1)$, $(1,0)$, $(1,1), \ldots$, $(a-1,b-1)$. Furthermore, the relative order of any pair of cards with the same labels is never altered. But this is exactly the same as an ab-unshuffle, if, at the beginning of such an unshuffle, we label each of the cards with one of the labels $(0,0)$, $(0,1)$, $\ldots$, $(0,b-1)$, $(1,0)$, $(1,1)$, $\ldots$, $(a-1,b-1)$. This completes the proof. □

In Figure 3.11, we show the labels for a 2-unshuffle of a deck with 10 cards. There are 4 cards with the label 0 and 6 cards with the label 1, so if the 2-unshuffle is performed, the first stack will have 4 cards and the second stack will have 6 cards. When this unshuffle is performed, the deck ends up in the identity ordering.

In Figure 3.12, we show the labels for a 4-unshuffle of the same deck (because there are four labels being used). This figure can also be regarded as an example of a pair of 2-unshuffles, as described in the proof above. The first 2-unshuffle will use the second coordinate of the labels to determine the stacks. In this case, the two stacks contain the cards whose values are

$$\{5,1,6,2,7\} \text{ and } \{8,9,3,4,10\} .$$

After this 2-unshuffle has been performed, the deck is in the order shown in Figure 3.11, as the reader should check. If we wish to perform a 4-unshuffle on the deck, using the labels shown, we sort the cards lexicographically, obtaining the four stacks

$$\{1,2\},\ \{3,4\},\ \{5,6,7\},\ \text{and } \{8,9,10\} .$$

When these stacks are combined, we once again obtain the identity ordering of the deck. The point of the above theorem is that both sorting procedures always lead to the same initial ordering.

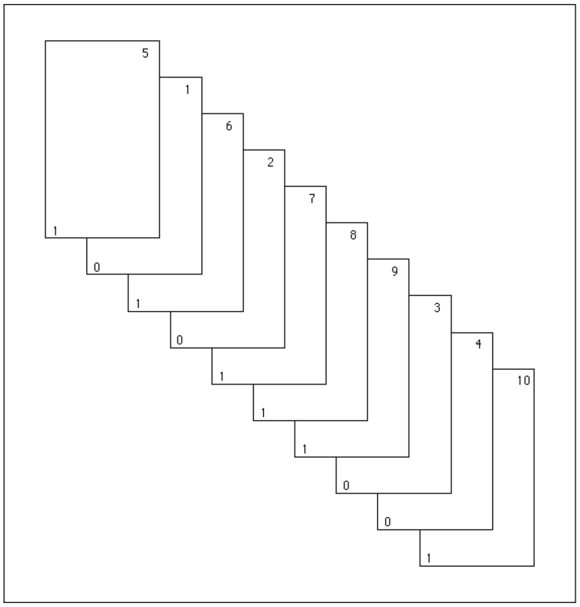

Figure 3.11: Before a 2-unshuffle.

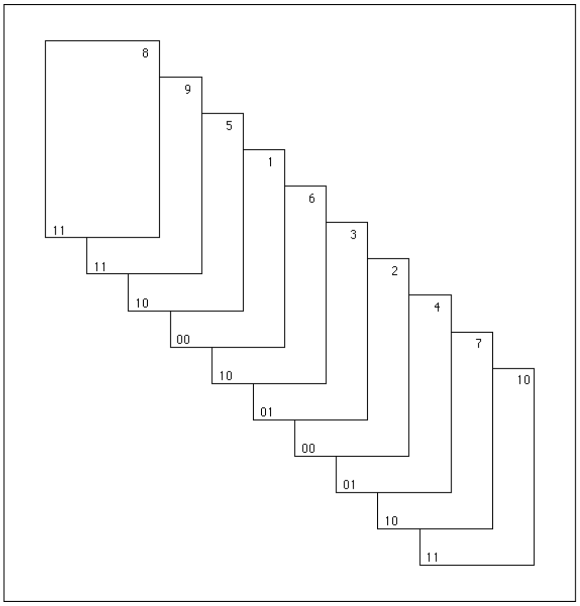

Figure 3.12: Before a 4-unshuffle.

Theorem 3.10 If D is any ordering that is the result of applying an a-shuffle and then a b-shuffle to the identity ordering, then the probability assigned to D by this pair of operations is the same as the probability assigned to D by the process of applying an ab-shuffle to the identity ordering.

Proof. Call the sample space of a-shuffles S_a. If we label the stacks by the integers from 0 to $a - 1$, then each cut-interleaving pair, i.e., shuffle, corresponds to exactly one n-digit base a integer, where the ith digit in the integer is the stack of which the ith card is a member. Thus, the number of cut-interleaving pairs is equal to the number of n-digit base a integers, which is a^n. Of course, not all of these pairs leads to different orderings. The number of pairs leading to a given ordering will be discussed later. For our purposes it is enough to point out that it is the cut-interleaving pairs that determine the probability assignment.

The previous theorem shows that there is a 1-1 correspondence between $S_{a,b}$ and S_{ab}. Furthermore, corresponding elements give the same ordering when applied to the identity ordering. Given any ordering D, let m_1 be the number of elements of $S_{a,b}$ which, when applied to the identity ordering, result in D. Let m_2 be the number of elements of S_{ab} which, when applied to the identity ordering, result in D. The previous theorem implies that $m_1 = m_2$. Thus, both sets assign the probability

$$\frac{m_1}{(ab)^n}$$

to D. This completes the proof. □

Connection with the Birthday Problem

There is another point that can be made concerning the labels given to the cards by the successive unshuffles. Suppose that we 2-unshuffle an n-card deck until the labels on the cards are all different. It is easy to see that this process produces each permutation with the same probability, i.e., this is a random process. To see this, note that if the labels become distinct on the sth 2-unshuffle, then one can think of this sequence of 2-unshuffles as one 2^s-unshuffle, in which all of the stacks determined by the unshuffle have at most one card in them (remember, the stacks correspond to the labels). If each stack has at most one card in it, then given any two cards in the deck, it is equally likely that the first card has a lower or a higher label than the second card. Thus, each possible ordering is equally likely to result from this 2^s-unshuffle.

Let T be the random variable that counts the number of 2-unshuffles until all labels are distinct. One can think of T as giving a measure of how long it takes in the unshuffling process until randomness is reached. Since shuffling and unshuffling are inverse processes, T also measures the number of shuffles necessary to achieve randomness. Suppose that we have an n-card deck, and we ask for $P(T \leq s)$. This equals $1 - P(T > s)$. But $T > s$ if and only if it is the case that not all of the labels after s 2-unshuffles are distinct. This is just the birthday problem; we are asking for the probability that at least two people have the same birthday, given that we have n people and there are 2^s possible birthdays. Using our formula from Example 3.3, we find that

$$P(T > s) = 1 - \binom{2^s}{n} \frac{n!}{2^{sn}} \ . \tag{3.4}$$

In Chapter 6, we will define the average value of a random variable. Using this idea, and the above equation, one can calculate the average value of the random variable T (see Exercise 6.1.41). For example, if $n = 52$, then the average value of T is about 11.7. This means that, on the average, about 12 riffle shuffles are needed for the process to be considered random.

Cut-Interleaving Pairs and Orderings

As was noted in the proof of Theorem 3.10, not all of the cut-interleaving pairs lead to different orderings. However, there is an easy formula which gives the number of such pairs that lead to a given ordering.

Theorem 3.11 If an ordering of length n has r rising sequences, then the number of cut-interleaving pairs under an a-shuffle of the identity ordering which lead to the ordering is

$$\binom{n + a - r}{n} \ .$$

Proof. To see why this is true, we need to count the number of ways in which the cut in an a-shuffle can be performed which will lead to a given ordering with r rising sequences. We can disregard the interleavings, since once a cut has been made, at most one interleaving will lead to a given ordering. Since the given ordering has r rising sequences, $r - 1$ of the division points in the cut are determined. The remaining $a - 1 - (r - 1) = a - r$ division points can be placed anywhere. The number of places to put these remaining division points is $n + 1$ (which is the number of spaces between the consecutive pairs of cards, including the positions at the beginning and the end of the deck). These places are chosen with repetition allowed, so the number of ways to make these choices is

$$\binom{n + a - r}{a - r} = \binom{n + a - r}{n} \ .$$

In particular, this means that if D is an ordering that is the result of applying an a-shuffle to the identity ordering, and if D has r rising sequences, then the probability assigned to D by this process is

$$\frac{\binom{n+a-r}{n}}{a^n} \ .$$

This completes the proof. □

The above theorem shows that the essential information about the probability assigned to an ordering under an a-shuffle is just the number of rising sequences in the ordering. Thus, if we determine the number of orderings which contain exactly r rising sequences, for each r between 1 and n, then we will have determined the distribution function of the random variable which consists of applying a random a-shuffle to the identity ordering.

The number of orderings of $\{1, 2, \ldots, n\}$ with r rising sequences is denoted by $A(n, r)$, and is called an Eulerian number. There are many ways to calculate the values of these numbers; the following theorem gives one recursive method which follows immediately from what we already know about a-shuffles.

Theorem 3.12 Let a and n be positive integers. Then

$$a^n = \sum_{r=1}^{a} \binom{n+a-r}{n} A(n, r) \ . \tag{3.5}$$

Thus,

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

In addition,

$$A(n, 1) = 1 \ .$$

Proof. The second equation can be used to calculate the values of the Eulerian numbers, and follows immediately from the Equation 3.5. The last equation is a consequence of the fact that the only ordering of $\{1, 2, \ldots, n\}$ with one rising sequence is the identity ordering. Thus, it remains to prove Equation 3.5. We will count the set of a-shuffles of a deck with n cards in two ways. First, we know that there are a^n such shuffles (this was noted in the proof of Theorem 3.10). But there are $A(n, r)$ orderings of $\{1, 2, \ldots, n\}$ with r rising sequences, and Theorem 3.11 states that for each such ordering, there are exactly

$$\binom{n+a-r}{n}$$

cut-interleaving pairs that lead to the ordering. Therefore, the right-hand side of Equation 3.5 counts the set of a-shuffles of an n-card deck. This completes the proof. □

Random Orderings and Random Processes

We now turn to the second question that was asked at the beginning of this section: What do we mean by a "random" ordering? It is somewhat misleading to think about a given ordering as being random or not random. If we want to choose a random ordering from the set of all orderings of $\{1, 2, \ldots, n\}$, we mean that we want every ordering to be chosen with the same probability, i.e., any ordering is as "random" as any other.

The word “random” should really be used to describe a process. We will say that a process that produces an object from a (finite) set of objects is a random process if each object in the set is produced with the same probability by the process. In the present situation, the objects are the orderings, and the process which produces these objects is the shuffling process. It is easy to see that no a-shuffle is really a random process, since if T_1 and T_2 are two orderings with a different number of rising sequences, then they are produced by an a-shuffle, applied to the identity ordering, with different probabilities.

Variation Distance

Instead of requiring that a sequence of shuffles yield a process which is random, we will define a measure that describes how far away a given process is from a random process. Let X be any process which produces an ordering of $\{1, 2, \ldots, n\}$. Define $f_X(\pi)$ be the probability that X produces the ordering π. (Thus, X can be thought of as a random variable with distribution function f.) Let Ω_n be the set of all orderings of $\{1, 2, \ldots, n\}$. Finally, let $u(\pi) = 1/|\Omega_n|$ for all $\pi \in \Omega_n$. The function u is the distribution function of a process which produces orderings and which is random. For each ordering $\pi \in \Omega_n$, the quantity

$$|f_X(\pi) - u(\pi)|$$

is the difference between the actual and desired probabilities that X produces π. If we sum this over all orderings π and call this sum S, we see that $S = 0$ if and only if X is random, and otherwise S is positive. It is easy to show that the maximum value of S is 2, so we will multiply the sum by 1/2 so that the value falls in the interval $[0, 1]$. Thus, we obtain the following sum as the formula for the *variation distance* between the two processes:

$$\parallel f_X - u \parallel = \frac{1}{2} \sum_{\pi \in \Omega_n} |f_X(\pi) - u(\pi)| \ .$$

Now we apply this idea to the case of shuffling. We let X be the process of s successive riffle shuffles applied to the identity ordering. We know that it is also possible to think of X as one 2^s-shuffle. We also know that f_X is constant on the set of all orderings with r rising sequences, where r is any positive integer. Finally, we know the value of f_X on an ordering with r rising sequences, and we know how many such orderings there are. Thus, in this specific case, we have

$$\parallel f_X - u \parallel = \frac{1}{2} \sum_{r=1}^{n} A(n, r) \left| \binom{2^s + n - r}{n} / 2^{ns} - \frac{1}{n!} \right| \ .$$

Since this sum has only n summands, it is easy to compute this for moderate sized values of n. For $n = 52$, we obtain the list of values given in Table 3.14.

To help in understanding these data, they are shown in graphical form in Figure 3.13. The program **VariationList** produces the data shown in both Table 3.14 and Figure 3.13. One sees that until 5 shuffles have occurred, the output of X is

Number of Riffle Shuffles	Variation Distance
1	1
2	1
3	1
4	0.9999995334
5	0.9237329294
6	0.6135495966
7	0.3340609995
8	0.1671586419
9	0.0854201934
10	0.0429455489
11	0.0215023760
12	0.0107548935
13	0.0053779101
14	0.0026890130

Table 3.14: Distance to the random process.

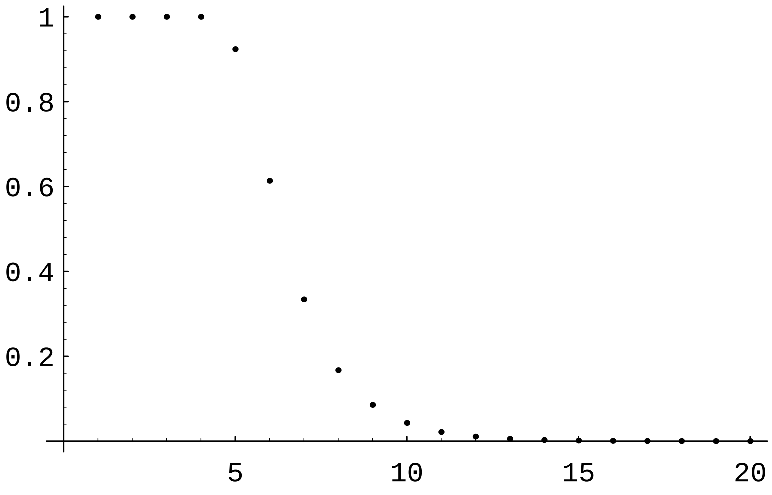

Figure 3.13: Distance to the random process.

very far from random. After 5 shuffles, the distance from the random process is essentially halved each time a shuffle occurs.

Given the distribution functions $f_X(\pi)$ and $u(\pi)$ as above, there is another way to view the variation distance $\parallel f_X - u \parallel$. Given any event T (which is a subset of S_n), we can calculate its probability under the process X and under the uniform process. For example, we can imagine that T represents the set of all permutations in which the first player in a 7-player poker game is dealt a straight flush (five consecutive cards in the same suit). It is interesting to consider how much the probability of this event after a certain number of shuffles differs from the probability of this event if all permutations are equally likely. This difference can be thought of as describing how close the process X is to the random process with respect to the event T.

Now consider the event T such that the absolute value of the difference between these two probabilities is as large as possible. It can be shown that this absolute value is the variation distance between the process X and the uniform process. (The reader is asked to prove this fact in Exercise 4.)

We have just seen that, for a deck of 52 cards, the variation distance between the 7-riffle shuffle process and the random process is about .334. It is of interest to find an event T such that the difference between the probabilities that the two processes produce T is close to .334. An event with this property can be described in terms of the game called New-Age Solitaire.

New-Age Solitaire

This game was invented by Peter Doyle. It is played with a standard 52-card deck. We deal the cards face up, one at a time, onto a discard pile. If an ace is encountered, say the ace of Hearts, we use it to start a Heart pile. Each suit pile must be built up in order, from ace to king, using only subsequently dealt cards. Once we have dealt all of the cards, we pick up the discard pile and continue. We define the Yin suits to be Hearts and Clubs, and the Yang suits to be Diamonds and Spades. The game ends when either both Yin suit piles have been completed, or both Yang suit piles have been completed. It is clear that if the ordering of the deck is produced by the random process, then the probability that the Yin suit piles are completed first is exactly 1/2.

Now suppose that we buy a new deck of cards, break the seal on the package, and riffle shuffle the deck 7 times. If one tries this, one finds that the Yin suits win about 75% of the time. This is 25% more than we would get if the deck were in truly random order. This deviation is reasonably close to the theoretical maximum of 33.4% obtained above.

Why do the Yin suits win so often? In a brand new deck of cards, the suits are in the following order, from top to bottom: ace through king of Hearts, ace through king of Clubs, king through ace of Diamonds, and king through ace of Spades. Note that if the cards were not shuffled at all, then the Yin suit piles would be completed on the first pass, before any Yang suit cards are even seen. If we were to continue playing the game until the Yang suit piles are completed, it would take 13 passes through the deck to do this. Thus, one can see that in a new deck, the Yin suits are in the most advantageous order and the Yang suits are in the least advantageous order. Under 7 riffle shuffles, the relative advantage of the Yin suits over the Yang suits is preserved to a certain extent.

Exercises

1 Given any ordering σ of $\{1, 2, \ldots, n\}$, we can define σ^{-1}, the inverse ordering of σ, to be the ordering in which the ith element is the position occupied by i in σ. For example, if $\sigma = (1, 3, 5, 2, 4, 7, 6)$, then $\sigma^{-1} = (1, 4, 2, 5, 3, 7, 6)$. (If one thinks of these orderings as permutations, then σ^{-1} is the inverse of σ.)

A *fall* occurs between two positions in an ordering if the left position is occupied by a larger number than the right position. It will be convenient to say that every ordering has a fall after the last position. In the above example, σ^{-1} has four falls. They occur after the second, fourth, sixth, and seventh positions. Prove that the number of rising sequences in an ordering σ equals the number of falls in σ^{-1}.

2 Show that if we start with the identity ordering of $\{1, 2, \ldots, n\}$, then the probability that an a-shuffle leads to an ordering with exactly r rising sequences equals

$$\frac{\binom{n+a-r}{n}}{a^n} A(n, r) \ ,$$

for $1 \le r \le a$.

3 Let D be a deck of n cards. We have seen that there are a^n a-shuffles of D. A coding of the set of a-unshuffles was given in the proof of Theorem 3.9. We will now give a coding of the a-shuffles which corresponds to the coding of the a-unshuffles. Let S be the set of all n-tuples of integers, each between 0 and $a - 1$. Let $M = (m_1, m_2, \ldots, m_n)$ be any element of S. Let n_i be the number of i's in M, for $0 \le i \le a - 1$. Suppose that we start with the deck in increasing order (i.e., the cards are numbered from 1 to n). We label the first n_0 cards with a 0, the next n_1 cards with a 1, etc. Then the a-shuffle corresponding to M is the shuffle which results in the ordering in which the cards labelled i are placed in the positions in M containing the label i. The cards with the same label are placed in these positions in increasing order of their numbers. For example, if $n = 6$ and $a = 3$, let $M = (1, 0, 2, 2, 0, 2)$. Then $n_0 = 2$, $n_1 = 1$, and $n_2 = 3$. So we label cards 1 and 2 with a 0, card 3 with a 1, and cards 4, 5, and 6 with a 2. Then cards 1 and 2 are placed in positions 2 and 5, card 3 is placed in position 1, and cards 4, 5, and 6 are placed in positions 3, 4, and 6, resulting in the ordering $(3, 1, 4, 5, 2, 6)$.

(a) Using this coding, show that the probability that in an a-shuffle, the first card (i.e., card number 1) moves to the ith position, is given by the following expression:

$$\frac{(a-1)^{i-1}a^{n-i} + (a-2)^{i-1}(a-1)^{n-i} + \cdots + 1^{i-1}2^{n-i}}{a^n} \ .$$

(b) Give an accurate estimate for the probability that in three riffle shuffles of a 52-card deck, the first card ends up in one of the first 26 positions. Using a computer, accurately estimate the probability of the same event after seven riffle shuffles.

4 Let X denote a particular process that produces elements of S_n, and let U denote the uniform process. Let the distribution functions of these processes be denoted by f_X and u, respectively. Show that the variation distance $\parallel f_X - u \parallel$ is equal to

$$\max_{T \subset S_n} \sum_{\pi \in T} \Big(f_X(\pi) - u(\pi) \Big) \ .$$

Hint: Write the permutations in S_n in decreasing order of the difference $f_X(\pi) - u(\pi)$.

5 Consider the process described in the text in which an n-card deck is repeatedly labelled and 2-unshuffled, in the manner described in the proof of Theorem 3.9. (See Figures 3.10 and 3.13.) The process continues until the labels are all different. Show that the process never terminates until at least $\lceil \log_2(n) \rceil$ unshuffles have been done.