(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. □

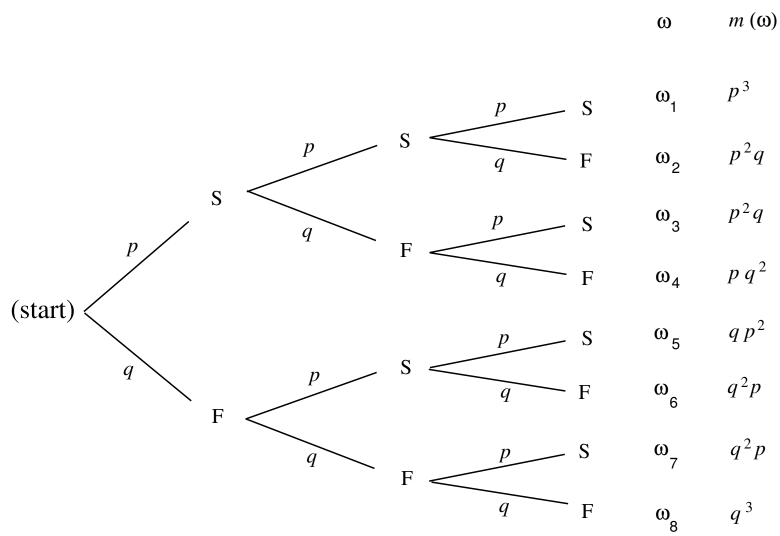

Figure 3.4: Tree diagram of three Bernoulli trials.

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 distribution 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 table 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,

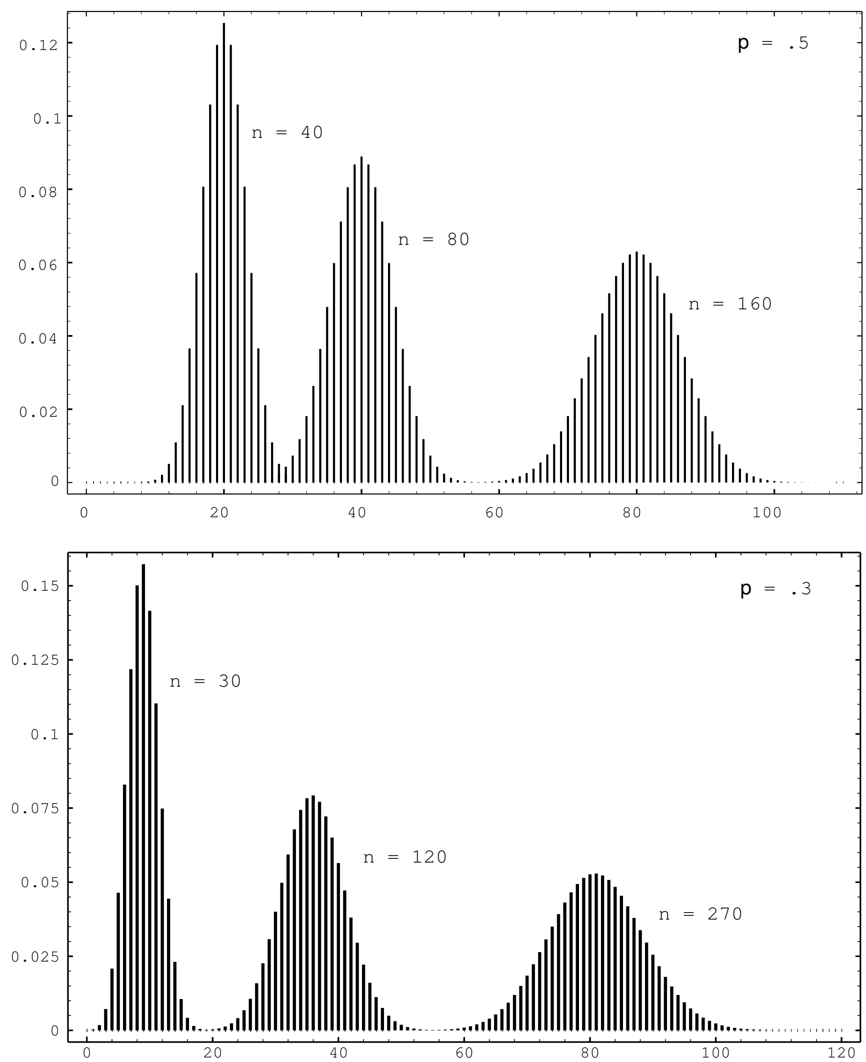

Figure 3.5: Binomial distributions.

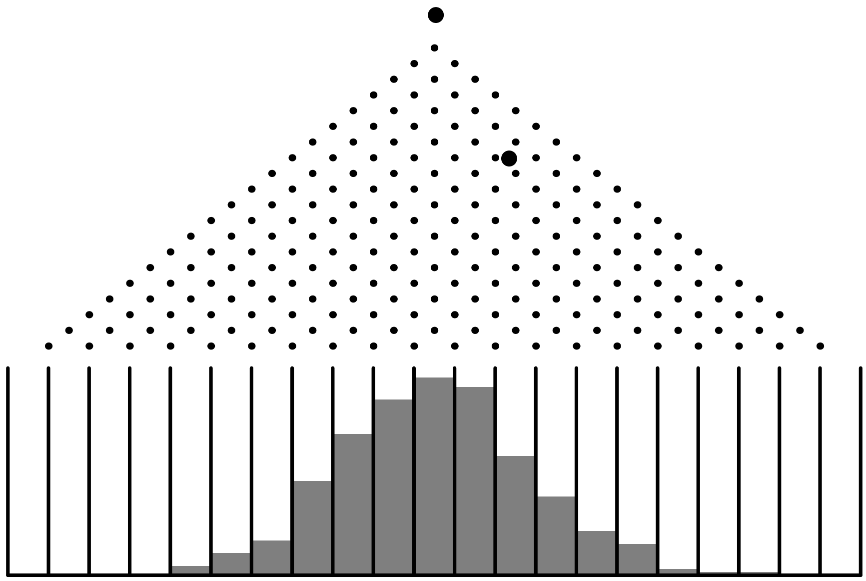

Figure 3.6: Simulation of the Galton board.

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

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.

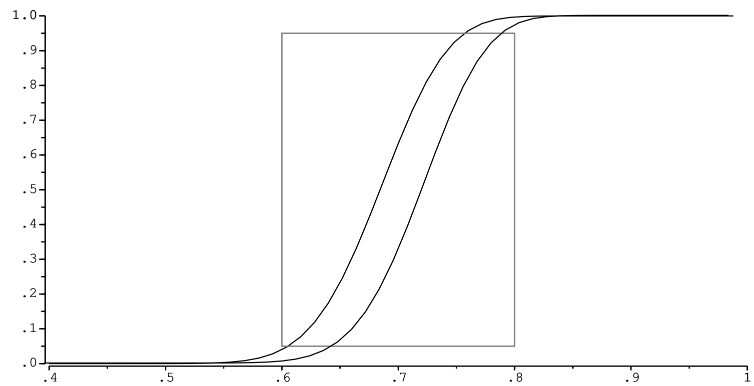

Figure 3.7: The power curve.

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 distribution 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. $\square$

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).