CHAPTER 3

DISCRETE DISTRIBUTIONS

In the sciences one often deals with "variables." Webster's dictionary defines a variable as a "quantity that may assume any one of a set of values." In statistics we deal with *random variables*—variables whose observed value is determined by chance. Many of the examples presented in previous chapters involved random variables even though the term was not used at the time. Random variables usually fall into one of two categories; they are either discrete or continuous. We begin by learning to recognize discrete random variables. The remainder of the chapter is devoted to the study of random variables of this type.

3.1 RANDOM VARIABLES

We begin by considering three examples, each of which involves a random variable. Random variables will be denoted by uppercase letters and their observed numerical values by lowercase letters.

Example 3.1.1. Consider the random variable X, the number of brown-eyed children born to a couple heterozygous for eye color. If the couple is assumed to have two children, a priori, before the fact, the variable X can assume any one of the values 0, 1, or 2. The variable is random in that brown eyes depend on the chance inheritance of a dominant gene at conception. If for a particular couple there are two brown-eyed children, we write $x = 2$.

Example 3.1.2. The basic premise underlying the field of immunology is that an animal is immunized by injection of a suitable antigen. In one study malignant plasmacytoma cells are exposed to lymphocytes carrying a specific antigen. It is hoped that these cells will fuse, because the fused cells retain the ability to grow continuously and also to retain the antibody characteristics of the antigen fused. In this way the animal is quickly immunized. Cells are exposed to the lymphocytes one at a time in the presence of polyethylene glycol, a fusion-promoting agent. It is known that the probability that such a cell will fuse is 1/2. Let Y denote the number of cells exposed to obtain the first fusion. The variable Y is random; a priori it can assume any value in the set $\{1, 2, 3, \ldots\}$. Recall from your study of calculus that a set such as this that consists of an infinite collection of isolated points is called a *countably infinite set.*

Example 3.1.3. In Example 1.1.2 we considered the variable T, the time at which the peak demand for electricity occurs per day. This variable is random, since its value is affected by such chance factors as time of the year, humidity, and temperature. It can conceivably assume any value in the 24-hour time span from 12 midnight one day to 12 midnight the next day.

It is easy to distinguish a discrete random variable from one that is not discrete. Just ask the question, "What are the possible values for the variable?" If the answer is a finite set or a countably infinite set, then the random variable is discrete; otherwise it is not. This idea leads to the following definition:

Definition 3.1.1 (Discrete random variable). A random variable is discrete if it can assume at most a finite or a countably infinite number of possible values.

The random variable X, the number of brown-eyed children in a two-child family, is discrete. Its set of possible values is the finite set $\{0, 1, 2\}$. The set $\{1, 2, 3, \ldots\}$ of possible values for Y, the number of cells exposed to obtain the first fusion of Example 3.1.2, is countably infinite. Thus Y is also a discrete random variable. The random variable T, the time of the peak demand for electricity at a power plant, is different from the others. Time is measured continuously, and T can conceivably assume any value in the interval $[0, 24)$, where 0 denotes 12 midnight one day and 24 denotes 12 midnight the next. This set of real numbers is neither finite nor countably infinite. Any time that you ask yourself the question, "What are the possible values for the random variable?" and are forced to admit that the set of possibilities includes some interval or continuous span of real numbers, then the random variable being studied is not discrete.

3.2 DISCRETE PROBABILITY DENSITIES

When dealing with a random variable, it is not enough just to determine what values are possible. We also need to determine what is probable. We must be able to predict in some sense the values that the variable is likely to assume at any time. Since the behavior of a random variable is governed by chance, these predictions must be made in the face of a great deal of uncertainty. The best that can be done is to describe the behavior of the random variable in terms of probabilities. Two functions are used to accomplish this. We shall refer to these as the *density function* and the *cumulative distribution function.* The former is known by a variety of names in the discrete case, some of the most commonly encountered ones being the probability function, the probability mass function, and the probability density function. In the discrete case, the density is denoted by either $p(x)$ or $f(x)$; in the continuous case it is almost always denoted by $f(x)$. For consistency we shall use $f(x)$ for the density in both cases. We begin by defining the density function for discrete random variables.

Definition 3.2.1 (Discrete density). Let X be a discrete random variable. The function f given by

$$f(x) = P[X = x]$$

for x real is called the density function for X.

There are several facts to note concerning the density in the discrete case. First, f is defined on the entire real line, and for any given real number x, $f(x)$ is the probability that the random variable X assumes the value x. For example, $f(2)$ is the probability that the random variable X assumes the numerical value of 2. Second, since $f(x)$ is a probability, $f(x) \geq 0$ regardless of the value of x. Third, if we sum f over all values of X that occur with nonzero probability, the sum must be 1. The following two conditions are necessary and sufficient conditions for a function f to be a discrete density. That is, if a function satisfies both of these conditions then it can be viewed as representing the density for some discrete random variable; if it fails to satisfy both then it cannot be the density for any discrete random variable:

Necessary and Sufficient Conditions for a Function to be a Discrete Density

1. $f(x) \geq 0$

2. $\sum_{\text{all } x} f(x) = 1$

The next example illustrates these ideas.

Example 3.2.1. Consider the random variable Y, the number of cells exposed to antigen-carrying lymphocytes in the presence of polyethylene glycol to obtain the first fusion (see Example 3.1.2). We know that under these conditions the probability that a given cell will fuse is 1/2. Thus the probability that it will not fuse is also 1/2. It is reasonable to assume that the cells behave independently. The possible values for Y are $\{1, 2, 3, \ldots\}$. The probability that the first cell will fuse is 1/2. That is,

$$P[Y = 1] = f(1) = 1/2$$

The probability that the first cell will not fuse but the second one will, yielding a value of 2 for Y, is

$$\begin{aligned} P[Y = 2] &= f(2) = P[\text{first cell does not fuse}]P[\text{second cell does fuse}] \\ &= 1/2 \cdot 1/2 = 1/4 \end{aligned}$$

Similarly,

$$P[Y = 3] = f(3) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$$

We can summarize the entire probability structure for Y in a density table (see Table 3.1). This is a table giving the possible values for the random variable in the first row and their corresponding probabilities in the second row. Note that there is an obvious pattern to the entries in row 2. When this occurs, we can find a closed-form expression for the density. In this case

$$f(y) = \begin{cases} (1/2)^y & y = 1, 2, 3, \ldots \\ 0 & \text{elsewhere} \end{cases}$$

Is this really a density? This function is obviously nonnegative, but does it sum to 1? To see this, note that

$$\sum_{\text{all } y} f(y) = \sum_{y=1}^{\infty} (1/2)^y$$

is a geometric series with first term $a = 1/2$ and common ratio $r = 1/2$. The properties of geometric series are well known. In particular, recall from elementary calculus that such a series can converge or diverge. The following fact will be useful in the material that follows:

Convergence of geometric series

Let $\sum_{k=1}^{\infty} ar^{k-1}$ be a geometric series.

The series converges to $\frac{a}{1-r}$ provided $|r| < 1$.

If we apply this result here, we see that

$$\sum_{y=1}^{\infty} (1/2)^y = \frac{a}{1-r} = \frac{1/2}{1-1/2} = 1$$

and the function f is a density.

Even though a discrete density is defined on the entire real line, it is only necessary to specify the density for those values y for which $f(y) \neq 0$. For instance, in the previous example we can write

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

It is understood that $f(y) = 0$ for all other real numbers.

Once it is known that a function is a density, it can be used to answer questions concerning the behavior of Y.

TABLE 3.1

y	1	2	3	$4 \cdots$
$P[Y = y] = f(y)$	1/2	$1/2 \cdot 1/2$	$1/2 \cdot 1/2 \cdot 1/2$	$1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdots$

Example 3.2.2. What is the probability that we will need to expose four or more cells to antigen-carrying lymphocytes in the presence of polyethylene glycol to obtain the first fusion? That is, what is $P[Y \geq 4]$? The density for Y is

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

Although the desired probability can be found directly, it is easier to use subtraction:

$$\begin{aligned} P[Y \geq 4] &= 1 - P[Y < 4] \\ &= 1 - P[Y \leq 3] \\ &= 1 - (P[Y = 1] + P[Y = 2] + P[Y = 3]) \\ &= 1 - (f(1) + f(2) + f(3)) \\ &= 1 - ((1/2)^1 + (1/2)^2 + (1/2)^3) \\ &= 1 - (1/2 + 1/4 + 1/8) \\ &= 1 - 7/8 = 1/8 \end{aligned}$$

Cumulative Distribution

The second function used to compute probabilities is the cumulative distribution function F. Most of the statistical tables used in the material that follows are tables of the cumulative distribution function for some pertinent random variable.

The word "cumulative" suggests the role of this function. It sums or accumulates the probabilities found by means of the density. This function is defined as follows:

Definition 3.2.2 (Cumulative distribution—discrete). Let X be a discrete random variable with density f. The cumulative distribution function for X, denoted by F, is defined by

$$F(x) = P[X \leq x] \qquad \text{for } x \text{ real}$$

Consider a specific real number x_0. To find $P[X \leq x_0] = F(x_0)$, we sum the density f over all values of X that occur with nonzero probability that are less than or equal to x_0. That is, computationally,

$$F(x_0) = \sum_{x \leq x_0} f(x)$$

This idea is illustrated in Example 3.2.3.

Example 3.2.3. Certain genes produce such a tremendous deviation from normal that the organism is unable to survive. Such genes are called lethal genes. An example is the gene that produces a yellow coat in mice, Y. This gene is dominant over that for gray, y. Normal genetic theory predicts that when two yellow mice heterozygous for this trait (Yy) mate, 1/4 of the offspring will be gray and 3/4 will be yellow. Biologists have observed that these predicted proportions do not, in fact, occur, but that the actual

TABLE 3.2

x	0	1	2	3
$P[X = x] = f(x)$	1/27	6/27	12/27	8/27

TABLE 3.3

x	0	1	2	3
$P[X \leq x] = F(x)$	1/27	7/27	19/27	27/27

TABLE 3.4

y	1	2	3	4 · · ·
$P[Y \leq y] = F(y)$	8/16	12/16	14/16	15/16 · · ·

percentages produced are 1/3 gray and 2/3 yellow. It has been established that this shift is caused by the fact that 1/4 of the embryos, those homozygous for yellow (YY), do not develop. This leaves only two genotypes, Yy and yy, occurring in a ratio of 2 to 1, with the former producing a mouse with a yellow coat. For this reason, the gene Y is said to be lethal.

The density for X, the number of yellow mice in a litter of size 3, is shown in Table 3.2, and its cumulative distribution is given in Table 3.3. Notice that

$$F(0) = P[X \leq 0] = P[X = 0] = 1/27$$
$$F(1) = P[X \leq 1] = P[X = 0] + P[X = 1] = 1/27 + 6/27$$
$$F(2) = P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2]$$
$$= 1/27 + 6/27 + 12/27$$
$$F(3) = P[X \leq 3] = 1$$

For discrete random variables that can assume only a finite number of possible values, the last entry in the bottom row of the cumulative table will always be 1.

Although cumulative probabilities are often given in table form as in the preceding example, it is sometimes possible to find express F in equation form. Example 3.2.4 illustrates this idea.

Example 3.2.4. Consider the random variable Y of Example 3.2.1 with density

$$f(y) = (1/2)^y \qquad y = 1, 2, 3, \ldots$$

A partial cumulative table for Y is shown in Table 3.4. It is formed by summing the probabilities given in the density table, Table 3.1. It is helpful to have a closed-form expression for F. In this case it is easy to obtain such an expression. By definition,

$$F(y_0) = \sum_{y \leq y_0} f(y)$$

If we let $[y_0]$ denote the greatest integer less than or equal to y_0, then in this case $F(y_0)$ can be expressed as

$$F(y_0) = \sum_{y=1}^{[y_0]} (1/2)^y$$
$$= \sum_{y=1}^{[y_0]} (1/2)(1/2)^{y-1}$$

Recall from elementary calculus that the sum of the first n terms of a geometric series is given by

Sum of first n terms: Geometric series

$$\sum_{k=1}^{n} ar^{k-1} = \frac{a(1-r^n)}{1-r} \qquad r \neq 1$$

where a is the first term of the series and r is the common ratio.

Apply this result with $a = 1/2$ and $r = 1/2$, to obtain

$$F(y_0) = \frac{(1/2)[1-(1/2)^{[y_0]}]}{1-1/2}$$
$$= 1 - (1/2)^{[y_0]}$$

The probability that at most seven cells must be exposed to obtain the first fusion is given by

$$P[Y \leq 7] = F(y) = 1 - (1/2)^7 = \frac{127}{128}$$

3.3 EXPECTATION AND DISTRIBUTION PARAMETERS

The density function of a random variable completely describes the behavior of the variable. However, associated with any random variable are constants, or "parameters," that are descriptive. Knowledge of the numerical values of these parameters gives the researcher quick insight into the nature of the variables. We consider three such parameters: the mean μ, the variance σ^2, and the standard deviation σ. If the exact density of the random variable is known, then the numerical value of each parameter can be found from mathematical considerations. That is the topic of this section. If the only thing available to the researcher is a set of observations on the random variable (a data set), then the values of these parameters cannot be found exactly. They must be approximated by using statistical techniques. That is the topic of much of the remainder of this text.

To understand the reasoning behind most statistical methods, it is necessary to become familiar with one general concept, namely, the idea of *mathematical expectation* or *expected value.* This concept is used in defining many statistical parameters and provides the logical basis for most of the methods of statistical inference presented later in this text.

A simple example will illustrate the basic idea of expectation. Consider the roll of a single fair die, and let X denote the number that is obtained. The possible values for X are 1, 2, 3, 4, 5, 6, and since the die is fair, the probability associated with each value is 1/6. The density for X is given by

$$f(x) = 1/6 \qquad x = 1, 2, 3, 4, 5, 6$$

When we ask for the expected value of X, we are asking for the *long-run theoretical average value of* X. If we imagine rolling the die over and over and recording the value of X for each roll, then we are asking for the theoretical average value of the rolls as the number of rolls approaches infinity. Since the density for X is symmetric and known, this average can be found intuitively. Notice that since $P[X = 1] = P[X = 6] = 1/6$, in the long run we expect to roll as many 1's as 6's. These values should counterbalance one another, and their average value is $(6 + 1)/2 = 3.5$. We also expect to roll as many 2's as 4's; these numbers also average to 3.5. Likewise, the numbers 3 and 4 are expected to counterbalance one another; they average 3.5. Logic dictates that, in the long run the average or expected value of X is 3.5. We write this as $E[X] = 3.5$. Notice that this value can be calculated from the density for X as follows:

$$E[X] = 1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 3.5$$

or

$$E[X] = \sum_{\text{all } x} (\text{value of } x)(\text{probability})$$

Of course, the characteristic that makes finding this expectation easy is the symmetry of the density. Can we develop a definition of expectation that will work for nonsymmetric densities and that will apply not only to X, but also to random variables that are functions of X? The answer is "yes," and the desired definition is given in Definition 3.3.1. Let us point out that in most problems interest centers first on $E[X]$. However, expectations for functions of X such as X^2, $(X - c)^2$, where c is a constant and e^{tX} are especially useful in statistical theory. For this reason, the definition of expected value is given in general terms. We now define what we mean by the expected value of some function of X which we denote by $H(X)$.

Definition 3.3.1 (Expected value). Let X be a discrete random variable with density f. Let $H(X)$ be a random variable. The expected value of $H(X)$, denoted by $E[H(X)]$, is given by

$$E[H(X)] = \sum_{\text{all } x} H(x)f(x)$$

provided $\sum_{\text{all } x} |H(x)| f(x)$ is finite. Summation is over all values of X that occur with nonzero probability.

Note that in the special case in which $H(X) = X$, we obtain the expected value of X from this definition. Thus we see that

Expected Value of X

$$E[X] = \sum_{\text{all } x} x f(x)$$

One other thing to note concerning this definition is the fact the restriction that $\sum_{\text{all } x} |H(x)| f(x)$ exists is not particularly restrictive in practice. If the set of possible values for X is finite, it *will* be satisfied; if the set of possible values for X is countably infinite, it will *usually* be satisfied. However, it is possible to concoct a density f and a function $H(X)$ for which the series $\sum_{\text{all } x} |H(x)| f(x)$ does not converge. (See Exercise 22.) In this case we say that the expected value of the random variable $H(X)$ does not exist. An example will illustrate the use of Definition 3.3.1. Please realize that the density has been greatly oversimplified for purposes of illustration!

Example 3.3.1. A drug is used to maintain a steady heart rate in patients who have suffered a mild heart attack. Let X denote the number of heartbeats per minute obtained per patient. Consider the hypothetical density given in Table 3.5. What is the average heart rate obtained by all patients receiving this drug? That is, what is $E[X]$? By Definition 3.3.1,

$$\begin{aligned} E[X] &= \sum_{\text{all } x} H(x) f(x) \\ &= \sum_{\text{all } x} x f(x) \\ &= 40(.01) + 60(.04) + 68(.05) + \cdots + 100(.01) \\ &= 70 \end{aligned}$$

Since the number of possible values for X is finite, $\sum_{\text{all } x} |x| f(x)$ exists. Thus we can say that the *average* heart rate obtained by patients using this drug is 70 heartbeats per minute. Intuitively, we should have expected this result. Notice the symmetry of the density. In the long run we would expect as many patients with heart rates of 100 as with heart rates of 40; as many with a rate of 60 as with a rate of 80. Similarly, the rates of 68 and 72 occur with the same frequency. Each of these pairs averages to 70, the value obtained by the remaining 80% of the patients. Common sense points to 70 as the expected value for X.

When used in a statistical context, the expected value of a random variable X is referred to as its *mean* and is denoted by μ or μ_X. That is, the terms *expected*

TABLE 3.5

x	40	60	68	70	72	80	100
$f(x)$	.01	.04	.05	.80	.05	.04	.01

value and *mean* are interchangeable, as are the symbols $E[X]$ and μ. The mean can be thought of as a measure of the "center of location" in the sense that it indicates where the "center" of the density lies. For this reason, the mean is often referred to as a "location" parameter. To emphasize these points, let us summarize the preceding discussion.

Notes on the Expected Value of a Random Variable X

1. The expected value of a random variable is its theoretical average value. It is denoted by $E[X]$ and can be calculated from knowledge of the density for X.
2. In a statistical setting, the average value of X is called its *mean value*. Hence the terms *average value, mean value,* and *expected value* are interchangeable.
3. The mean value of X is denoted by the Greek symbol μ (*mu*). Hence the symbols μ and $E[X]$ are interchangeable.
4. The mean or expected value of X is one measure of the location of the center of the X values. For this reason, μ is called a "location" parameter.

There are three rules for handling expected values that are useful in justifying statistical procedures in later chapters. These rules hold for both continuous and discrete random variables. The rules are stated and illustrated here. We outline the proofs of the first two as exercises; the proof of rule 3 must be deferred until Chap. 5.

Theorem 3.3.1 (Rules for expectation). Let X and Y be random variables and let c be any real number.

1. $E[c] = c$ (The expected value of any constant is that constant.)
2. $E[cX] = cE[X]$ (Constants can be factored from expectations.)
3. $E[X + Y] = E[X] + E[Y]$ (The expected value of a sum is equal to the sum of the expected values.)

Example 3.3.2. Let X and Y be random variables with $E[X] = 7$ and $E[Y] = -5$. Then

$$\begin{aligned} E[4X - 2Y + 6] &= E[4X] + E[-2Y] + E[6] && \text{Rule 3} \\ &= 4E[X] + (-2)E[Y] + E[6] && \text{Rule 2} \\ &= 4E[X] - 2E[Y] + 6 && \text{Rule 1} \\ &= 4(7) - 2(-5) + 6 \\ &= 44 \end{aligned}$$

Variance and Standard Deviation

Knowledge of the mean of a random variable is important, but this knowledge *alone* can be misleading. The next example should show you the problem.

Example 3.3.3. Suppose that we wish to compare a new drug to that of Example 3.3.1. Let X denote the number of heartbeats per minute obtained using the old drug and Y the number per minute obtained with the new drug. The hypothetical density of

TABLE 3.6

x	40	60	68	70	72	80	100
$f(x)$	.01	.04	.05	.80	.05	.04	.01

y	40	60	68	70	72	80	100
$f(y)$	.40	.05	.04	.02	.04	.05	.40

each of these variables is given in Table 3.6. Since each of the densities is symmetric, inspection shows that $\mu_X = \mu_Y = 70$. Each drug produces *on the average* the same number of heartbeats per minute. However, there is obviously a drastic difference between the two drugs that is not being detected by the mean. The old drug produces fairly consistent reactions in patients, with 90% differing from the mean by at most 2; very few (2%) have an extreme reaction to the drug. However, the new drug produces highly diverse responses. Only 10% of the patients have heart rates within 2 units of the mean, whereas 80% show an extreme reaction. If we examined only the mean, we would conclude that the two drugs had identical effects—but nothing could be further from the truth!

It is obvious from Example 3.3.3 that something is not being measured by the mean. That something is *variability*. We must find a parameter that reflects consistency or the lack of it. We want the measure to assume a large positive value if the random variable fluctuates in the sense that it often assumes values far from its mean; the measure should assume a small positive value if the values of X tend to cluster closely about the mean. There are several ways to define such a measure. The most widely used is the *variance*.

Definition 3.3.2 (Variance). Let X be a random variable with mean μ. The variance of X, denoted by Var X, or σ^2, is given by

$$\text{Var } X = \sigma^2 = E[(X - \mu)^2]$$

Note that the variance measures variability by considering $X - \mu$, the difference between the variable and its mean. The difference is squared so that negative values will not cancel positive ones in the process of finding the expected value. When expressed in the form $E[(X - \mu)^2]$, it is easy to see that σ^2 has the properties that we want. When the variable X often assumes values far from μ, σ^2 will be a large positive number; when the values of X tend to fall close to μ, σ^2 will assume a small positive value. Figure 3.1 illustrates the idea.

Usually, the definition of σ^2 is not used to compute the variance. Rather, we use an alternative form which is given in the following theorem.

Theorem 3.3.2 (Computational formula for σ^2)

$$\sigma^2 = \text{Var } X = E[X^2] - (E[X])^2$$

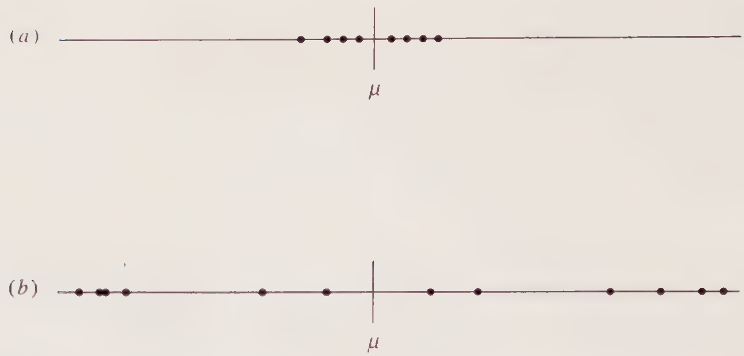

FIGURE 3.1
(*a*) A distribution with a small variance. Most of the data points, denoted by dots, lie fairly close to the average value, μ. Hence most of the differences, $x - \mu$, will be small; (*b*) a distribution with a large variance. Many of the data points lie far from the average value, μ.

Proof. By definition

$$\begin{aligned}\operatorname{Var} X &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2]\end{aligned}$$

Using the rules of expectation, Theorem 3.3.1, we obtain

$$\operatorname{Var} X = E[X^2] - 2\mu E[X] + \mu^2$$

Since the symbols μ and $E[X]$ are interchangeable,

$$\begin{aligned}\operatorname{Var} X &= E[X^2] - 2(E[X])^2 + (E[X])^2 \\ &= E[X^2] - (E[X])^2\end{aligned}$$

We illustrate the theorem by computing the variance of each of the random variables of Example 3.3.3.

Example 3.3.4. To find σ_X^2 and σ_Y^2 for the variables of Example 3.3.3, we first use Table 3.6 to find $E[X^2]$ and $E[Y^2]$. We know that $E[X] = E[Y] = 70$.

$$\begin{aligned}E[X^2] &= \sum_{\text{all } x} x^2 f(x) \\ &= (40^2)(.01) + (60^2)(.04) + \cdots + (100^2)(.01) \\ &= 4926.4\end{aligned}$$

$$\begin{aligned}E[Y^2] &= \sum_{\text{all } y} y^2 f(y) \\ &= (40^2)(.40) + (60^2)(.05) + \cdots + (100^2)(.40) \\ &= 5630.32\end{aligned}$$

By Theorem 3.3.2,

$$\begin{aligned}\operatorname{Var} X &= E[X^2] - (E[X])^2 \\ &= 4926.4 - 70^2 = 26.4 \\ \operatorname{Var} Y &= E[Y^2] - (E[Y])^2 \\ &= 5630.32 - 70^2 = 730.32\end{aligned}$$

As expected, Var $Y >$ Var X. Even though the drugs produce the same mean number of heartbeats per minute, they do not behave in the same way. The new drug is not as consistent in its effect as the old.

Note that the variance of a random variable reported alone is not very informative. Is a variance of 26.4 large or small? Only when this value is compared to the variance of a similar variable does it take on meaning. Hence variances are used often for comparative purposes to choose between two variables that otherwise appear to be identical. Also, note that the variance of a random variable is essentially a pure number whose associated units are often physically meaningless. When this occurs, the unit can be omitted. For example, the unit associated with the variance of Example 3.3.4 is a "squared heartbeat." This makes little sense, so in this case variance can be reported with no unit attached. To overcome this problem, a second measure of variability is employed. This measure is the nonnegative square root of the variance, and it is called the *standard deviation.* It has the advantage of having associated with it the same units as the original data.

Definition 3.3.3 (Standard deviation). Let X be a random variable with variance σ^2. The standard deviation of X, denoted by σ, is given by

$$\sigma = \sqrt{\operatorname{Var} X} = \sqrt{\sigma^2}$$

Example 3.3.5. The standard deviations of variables X and Y of Example 3.3.4 are, respectively,

$$\sigma_X = \sqrt{\operatorname{Var} X} = \sqrt{26.4} = 5.14 \text{ heartbeats per minute}$$
$$\sigma_Y = \sqrt{\operatorname{Var} Y} = \sqrt{730.32} = 27.02 \text{ heartbeats per minute}$$

To emphasize these points we present a brief summary of the important aspects of the standard deviation of a random variable X.

Properties of standard deviation

1. The standard deviation of X is defined as the nonnegative square root of its variance.
2. The standard deviation is denoted by σ, and the variance of X is denoted by σ^2.
3. A large standard deviation implies that the random variable X is rather inconsistent and somewhat hard to predict; a small standard deviation is an indication of consistency and stability.

4. Standard deviation is always reported in physical measurement units that match the original data. Variance is often unitless.

Just as there are three rules for expectation that help in simplifying complex expressions, so are there three rules for variance. These rules parallel those for expectation. Rules 1 and 2 can be proved by using the rules for expectation (see Exercise 20). The proof of rule 3 must be deferred until the notion of "independent random variables" has been formalized.

Theorem 3.3.3 (Rules for variance). Let X and Y be random variables and c any real number. Then

1. $\operatorname{Var} c = 0$
2. $\operatorname{Var} cX = c^2 \operatorname{Var} X$
3. If X and Y are independent, then $\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y$

(Two variables are independent if knowledge of the value assumed by one gives no clue to the value assumed by the other.)

Example 3.3.6. Let X and Y be independent with $\sigma_X^2 = 9$ and $\sigma_Y^2 = 3$. Then

$$\begin{aligned}
\operatorname{Var}[4X - 2Y + 6] &= \operatorname{Var}[4X] + \operatorname{Var}[-2Y] + \operatorname{Var} 6 && \text{Rule 3}\\
&= 16 \operatorname{Var} X + 4 \operatorname{Var} Y + \operatorname{Var} 6 && \text{Rule 2}\\
&= 16 \operatorname{Var} X + 4 \operatorname{Var} Y + 0 && \text{Rule 1}\\
&= 16(9) + 4(3) = 156
\end{aligned}$$

In this section we discussed three *theoretical* parameters associated with a random variable X. We showed not only how to determine their numerical values from knowledge of the density, but also how to interpret them physically. Keep these things in mind, for they play a major role in the study of statistical methods for analyzing experimental data.

3.4 GEOMETRIC DISTRIBUTION AND THE MOMENT GENERATING FUNCTION

In this section we consider two important topics. We introduce the first *family* of discrete random variables to be discussed in this text. Random variables are members of a family in the sense that each member of the family is characterized by a density function of the same mathematical form, differing only with respect to the numerical value of some pertinent parameter or parameters. This first family, called *geometric*, is used extensively in the areas of games of chance and in statistical quality control. It is named geometric because, as you will see, its theoretical properties are derived by applying the mathematical properties of the geometric series that you encountered in elementary calculus. The second topic is a discussion of the *moment generating function*. This is a function, derived from the density, that allows one to calculate ordinary moments of a distribution easily. This in turn makes it possible to calculate the mean and variance of a random variable without having to use the definitions of these terms to do so. In many cases, this approach is much simpler than a direct calculation from the definition. The function also provides a fingerprint or a unique identifier for each distribution. This idea will be illustrated later in this section.

Geometric Distribution

We begin by considering the family of *geometric* random variables. As you shall see, you have already encountered some random variables of this type even though the name "geometric random variable" was not mentioned at the time.

Geometric random variables arise in practice in experiments characterized by the following properties:

Geometric properties

1. The experiment consists of a series of trials. The outcome of each trial can be classed as being either a "success" (s) or a "failure" (f). A trial with this property is called a *Bernoulli* trial.
2. The trials are identical and independent in the sense that the outcome of one trial has no effect on the outcome of any other. The probability of success, p, remains the same from trial to trial.
3. The random variable X denotes the number of trials needed to obtain the first success.

The sample space for an experiment such as that just described is

$$S = \{s, fs, ffs, fffs, \ldots\}$$

Since the random variable X denotes the number of trials needed to obtain the first success, X assumes the values 1, 2, 3, 4, ... To find the density for X, we look for a pattern. Note that

$$P[X = 1] = P[\text{success on first trial}] = p$$
$$P[X = 2] = P[\text{fail on first trial and succeed on second trial}]$$

Since the trials are independent, the latter probability can be found by multiplying. That is,

$$\begin{aligned} P[X = 2] &= P\,[\text{fail on first trial and succeed on second trial}] \\ &= P[\text{fail on first trial}]P[\text{succeed on second trial}] \\ &= (1 - p)(p) \end{aligned}$$

Similarly,

$$\begin{aligned} P[X = 3] &= P[\text{fail on first trial and fail on second trial and succeed on third trial}] \\ &= (1 - p)(1 - p)(p) = (1 - p)^2 p \end{aligned}$$

TABLE 3.7

x	1	2	3	4	5	...
$f(x)$	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4p$	...

You should be able to see that the density for X is given by Table 3.7, where the probabilities given in row 2 of the table exhibit a definite pattern. This pattern can be expressed in closed form as

$$f(x) = (1-p)^{x-1}p \qquad x = 1, 2, 3, \ldots$$

We now define a geometric random variable as being any random variable with a density of this form.

Definition 3.4.1 (Geometric distribution). A random variable X is said to have a geometric distribution with parameter p if its density f is given by

$$f(x) = (1-p)^{x-1}p \qquad 0 < p < 1$$
$$x = 1, 2, 3, \ldots$$

The function f given in this definition is a density. It is obviously nonnegative. Furthermore,

$$\sum_{x=1}^{\infty} (1-p)^{x-1}p$$

is a geometric series with first term $a = p$ and common ratio $r = (1-p)$. Thus the series sums to

$$\frac{a}{1-r} = \frac{p}{1-(1-p)} = 1$$

as desired. From this argument the reason for the name "geometric" distribution should be apparent.

In Exercise 26 you are asked to verify that the general expression for the cumulative distributions function for a geometric random variable is

$$\boxed{F(x) = 1 - q^{[x]}}$$

where q is the probability of failure and $[x]$ is the greatest integer less than or equal to x.

Example 3.4.1. Random digits are integers selected from among {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} one at a time in such a way that at each stage in the selection process the integer chosen is just as likely to be one digit as any other. In simulation experiments it is often necessary to generate a series of random digits. This can be done in a number of ways, the most common being by means of a computerized random number generator.

In generating such a series, let X denote the number of trials needed to obtain the first zero. This experiment consists of a series of independent, identical trials with "success" being the generation of a zero. The probability of success is $p = 1/10$. Since X denotes the number of trials needed to obtain the first success, X is a geometric random variable. Its density is found by substituting the value 1/10 for p in the expression for f given in Definition 3.4.1. That is,

$$f(x) = (1 - p)^{x-1}p \qquad x = 1, 2, 3, \ldots$$

or

$$f(x) = (9/10)^{x-1}1/10 \qquad x = 1, 2, 3, \ldots$$

The cumulative distribution function for X is given by

$$F(x) = 1 - (.9)^{[x]}$$

Finding the mean of a geometric random variable from the definition is tricky! Consider the next example.

Example 3.4.2. Let us find the mean of the random variable X, the number of trials needed to obtain a zero when generating a series of random digits. By Definition 3.3.1,

$$\mu = E[X] = \sum_{x=1}^{\infty} xf(x)$$
$$= \sum_{x=1}^{\infty} x(9/10)^{x-1}1/10$$

That is,

$$E[X] = 1/10 + 18/100 + 243/1000 + 2916/10{,}000 + \cdots$$

This series is not geometric. Consider the series $(9/10)E[X]$.

$$(9/10)E[X] = 9/100 + 162/1000 + 2187/10{,}000 + 26{,}244/100{,}000 + \cdots$$

Subtracting the latter from the former, we obtain

$$(1/10)E[X] = 1/10 + 9/100 + 81/1000 + 729/10{,}000 + \cdots$$

This series is geometric with first term 1/10 and common ratio 9/10. Thus

$$(1/10)E[X] = \frac{1/10}{1 - 9/10} = 1$$

or

$$E[X] = \frac{1}{1/10} = 10$$

Moment Generating Function

As we have seen, the two expectations $E[X]$ and $E[X^2]$ are very useful, as they allow us to find the mean and variance of the random variable. These, and other expectations of the form $E[X^k]$ for k a positive integer, are examples of what are called *ordinary moments.* This term is defined as follows:

Definition 3.4.2 (Ordinary moments). Let X be a random variable. The k^{th} ordinary moment for X is defined as $E[X^k]$.

Thus $E[X] = \mu$ is the first ordinary moment for X; $E[X^2]$ is its second ordinary moment. The preceding example shows that finding ordinary moments, even the first moment, from the definition of expectation is not always easy. Fortunately, it is often possible to obtain a function, called the *moment generating function,* which will enable us to find these moments with less effort.

Definition 3.4.3 (Moment generating function). Let X be a random variable with density f. The moment generating function for X (m.g.f.) is denoted by $m_X(t)$ and is given by

$$m_X(t) = E[e^{tX}]$$

provided this expectation is finite for all real numbers t in some open interval $(-h, h)$.

Since each geometric random variable has a density of the same general form, it is possible to find a general expression for the moment generating function for such a variable. This expression is given in Theorem 3.4.1.

Theorem 3.4.1 (Geometric moment generating function). Let X be a geometric random variable with parameter p. The moment generating function for X is given by

$$m_X(t) = \frac{pe^t}{1 - qe^t} \qquad t < -\ln q$$

where $q = 1 - p$.

Proof. The density for X is given by

$$f(x) = q^{x-1}p \qquad x = 1, 2, 3, \ldots$$

By definition

$$\begin{aligned} m_X &= E[e^{tX}] \\ &= \sum_{\text{all } x} e^{tx} f(x) \\ &= \sum_{x=1}^{\infty} e^{tx} q^{x-1} p \\ &= pq^{-1} \sum_{x=1}^{\infty} (qe^t)^x \end{aligned}$$

The series on the right is a geometric series with first term qe^t and common ratio qe^t. Thus

$$\begin{aligned} m_X(t) &= pq^{-1}\left(\frac{qe^t}{1-qe^t}\right) \\ &= \frac{pe^t}{1-qe^t} \end{aligned}$$

provided $|r| = |qe^t| < 1$. Since the exponential function is nonnegative and $0 < q < 1$, this restriction implies that $qe^t < 1$. The inequality is solved for t as follows:

$$\begin{aligned} qe^t &< 1 \\ e^t &< 1/q \\ \ln e^t &< \ln 1/q \\ t &< \ln 1 - \ln q \\ t &< -\ln q \end{aligned}$$

The next theorem shows how the moment generating function can be used to generate ordinary moments for a random variable X. Its proof is based on the Maclaurin series expansion for e^z. Recall that this series is as follows:

Maclaurin Series Expansion for e^z

$$e^z = 1 + z + z^2/2! + z^3/3! + z^4/4! + \cdots$$

Theorem 3.4.2. Let $m_X(t)$ be the moment generating function for a random variable X. Then

$$\left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0} = E[X^k]$$

Proof. To prove this theorem, let $z = tX$. The Maclaurin series expansion for e^{tX} is

$$e^{tX} = 1 + tX + (tX)^2/2! + (tX)^3/3! + (tX)^4/4! + \cdots$$

By taking the expected value of each side of this equation, we obtain

$$\begin{aligned} m_X(t) = E[e^{tX}] &= E[1 + tX + t^2X^2/2! + t^3X^3/3! + t^4X^4/4! + \cdots] \\ &= 1 + tE[X] + t^2/2!E[X^2] + t^3/3!E[X^3] + t^4/4!E[X^4] + \cdots \end{aligned}$$

Differentiating this series term by term with respect to t, we see that

$$\frac{dm_X(t)}{dt} = E[X] + tE[X^2] + t^2/2!E[X^3] + t^3/3!E[X^4] + \cdots$$

When this derivative is evaluated at $t = 0$, every term except the first becomes 0. Hence

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = E[X]$$

Taking the second derivative of $m_X(t)$, we obtain

$$\frac{d^2m_X(t)}{dt^2} = E[X^2] + tE[X^3] + t^2/2!E[X^4] + \cdots$$

Evaluating this derivative at $t = 0$ yields

$$\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = E[X^2]$$

This procedure can be continued to show that

$$\left.\frac{d^km_X(t)}{dt^k}\right|_{t=0} = E[X^k]$$

for any positive integer k as desired.

Let us use the moment generating function to find a general expression for the mean and variance of a geometric distribution with parameter p.

Theorem 3.4.3. Let X be a geometric random variable with parameter p. Then

$$E[X] = 1/p \qquad \text{and} \qquad \operatorname{Var} X = q/p^2$$

Proof. For a geometric random variable with parameter p

$$\begin{aligned} m_X(t) &= \frac{pe^t}{1 - qe^t} \\ \frac{dm_X(t)}{dt} &= \frac{(1 - qe^t)pe^t + pe^tqe^t}{(1 - qe^t)^2} \\ &= \frac{pe^t}{(1 - qe^t)^2} \end{aligned}$$

Evaluating this derivative at $t = 0$, we obtain

$$\begin{aligned} E[X] &= \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{p}{(1 - q)^2} \\ &= p/p^2 \\ &= 1/p \end{aligned}$$

Taking the second derivative of $m_X(t)$, we obtain

$$\begin{aligned} \frac{d^2m_X(t)}{dt^2} &= \frac{(1 - qe^t)^2pe^t + 2pe^t(1 - qe^t)qe^t}{(1 - qe^t)^4} \\ &= \frac{pe^t(1 - qe^t)[(1 - qe^t) + 2qe^t]}{(1 - qe^t)^4} \\ &= \frac{pe^t(1 + qe^t)}{(1 - qe^t)^3} \end{aligned}$$

Evaluating this derivative at $t = 0$, we see that

$$E[X^2] = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{p(1+q)}{(1-q)^3} = \frac{(1+q)}{p^2}$$

Now

$$\begin{aligned}\operatorname{Var} X &= E[X^2] - (E[X])^2 \\ &= \frac{1+q}{p^2} - \frac{1}{p^2} \\ &= \frac{q}{p^2}\end{aligned}$$

We illustrate the use of these theorems by finding the moment generating function, mean, and variance for the random variable of Example 3.4.1.

Example 3.4.3. Consider the random variable X, the number of trials needed to obtain the first zero when generating a series of random digits. Since this random variable is geometric with parameter $p = 1/10$,

$$\begin{aligned}m_X(t) &= \frac{pe^t}{1-qe^t} = \frac{(1/10)e^t}{1-(9/10)e^t} \\ \mu &= E[X] = 1/p = 10 \\ \sigma^2 &= \operatorname{Var} X = q/p^2 = \frac{9/10}{(1/10)^2} = 90\end{aligned}$$

Note that this value for μ agrees with that obtained in Example 3.4.2.

The importance of the moment generating function for a random variable is not completely evident at this time. It does give us a way to find general expressions for the mean and variance as well as for the ordinary moments of an entire family of random variables. As we shall see later, the moment generating function, when it exists, serves as a fingerprint that completely identifies the random variable under study. That is, if a distribution has a moment generating function then it is unique. Thus, to identify a distribution from its moment generating function we need only look for and recognize a pattern and then the distribution is evident. For example, if an unknown random variable has moment generating function

$$m_X(t) \frac{.4e^t}{1-.6e^t}$$

then we know that the random variable follows a geometric distribution with $p = .4$, because the moment generating function assumes the general form

$$\frac{pe^t}{1-qe^t}$$

which is the geometric fingerprint.

3.5 BINOMIAL DISTRIBUTION

The next distribution to be studied is the *binomial* distribution. Once again, you have already seen some binomial random variables even though they were not labeled as such at the time. The theoretical basis for working with this distribution is the binomial theorem presented in most beginning algebra courses. The statement of this theorem is as follows:

Binomial theorem

For any two real numbers a and b and any positive integer n,

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

where $\binom{n}{k}$ is given by $\dfrac{n!}{k!(n-k)!}$.

To recognize a situation that involves a binomial random variable, you must be familiar with the assumptions that underlie this distribution, which are as follows:

Binomial properties

1. The experiment consists of a *fixed* number, n, of Bernoulli trials, trials that result in either a "success" (s) or a "failure" (f).
2. The trials are identical and independent, and therefore the probability of success, p, remains the same from trial to trial.
3. The random variable X denotes the number of successes obtained in the n trials.

Once we realize that the binomial model is appropriate from the physical description of the experiment, we shall want to describe the behavior of the binomial random variable involved. To do so, we need to consider the density for the random variable. To get an idea of the general form for the binomial density, let us consider the case in which $n = 3$. The sample space for such an experiment is

$$S = \{fff, sff, fsf, ffs, ssf, sfs, fss, sss\}$$

Since the trials are independent, the probability assigned to each sample point is found by multiplying. For example, the probabilities assigned to the sample points *fff* and *sff* are $(1-p)(1-p)(1-p) = (1-p)^3$ and $p(1-p)(1-p) = p(1-p)^2$, respectively. The random variable X assumes the value 0 only if the experiment results in the outcome *fff*. That is,

$$P[X = 0] = (1-p)^3$$

However, X assumes the value 1 if the experiment results in any one of the outcomes *sff*, *fsf*, or *ffs*. Thus

$$P[X = 1] = 3 \cdot p(1-p)^2$$

Similarly,

$$P[X = 2] = 3 \cdot p^2(1-p)$$

and

$$P[X = 3] = p^3$$

It is evident that for $x = 0, 1, 2, 3$

$$P[X = x] = c(x)p^x(1 - p)^{3-x}$$

where $c(x)$ denotes the number of sample points that correspond to x successes. Such a sample point is expressed as a permutation of three letters, with x of these being s's and the rest, $3 - x$, of these being f's. Using the formula for the number of permutations of indistinguishable objects studied in Chap. 1, we see that

$$c(x) = \frac{3!}{x!(3-x)!} = \binom{3}{x}$$

Thus the density for this binomial random variable is given by

$$f(x) = \binom{3}{x}p^x(1-p)^{3-x} \qquad x = 0, 1, 2, 3$$

To generalize this idea to n trials, we replace 3 by n to obtain the expression

$$f(x) = \binom{n}{x}p^x(1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

This suggests the formal definition of the binomial distribution.

Definition 3.5.1 (Binomial distribution). A random variable X has a binomial distribution with parameters n and p if its density is given by

$$f(x) = \binom{n}{x}p^x(1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$
$$0 < p < 1$$

where n is a positive integer.

To see that the function given in this definition is a density, note that it is nonnegative. Furthermore, by applying the binomial theorem with $k = x$, $a = p$, and $b = 1 - p$ it can be seen that

$$\sum_{x=0}^{n} \binom{n}{x}p^x(1-p)^{n-x} = [p + (1-p)]^n = 1$$

as desired.

Example 3.5.1. Recent studies of German air traffic controllers have shown that it is difficult to maintain accuracy when working for long periods of time on data display screens. A surprising aspect of the study is that the ability to detect spots on a radar screen decreases as their appearance becomes too rare. The probability of correctly identifying a signal is approximately .9 when 100 signals arrive per 30-minute period. This probability drops to .5 when only 10 signals arrive at random over a 30-minute period. The hypothesis is that unstimulated minds tend to wander. Let X denote the number of signals correctly identified in a 30-minute time span in which 10 signals arrive. This experiment consists of a series of $n = 10$ independent and identical Bernoulli trials with "success" being the correct identification of a signal. The probability of success is $p = 1/2$. Since X denotes the number of successes in a fixed number of trials, X is binomial. Its density is found by letting $n = 10$ and $p = 1/2$ in the expression for f given in Definition 3.5.1. That is,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n$$

or

$$f(x) = \binom{10}{x} (1/2)^x (1/2)^{10-x} \qquad x = 0, 1, 2, \ldots, 10$$

The next theorem summarizes other theoretical properties of the binomial distribution. Its proof is left as an exercise (Exercise 43).

Theorem 3.5.1. Let X be a binomial random variable with parameters n and p.

1. The moment generating function for X is given by

$$m_X(t) = (q + pe^t)^n \qquad q = 1 - p$$

2. $E[X] = \mu = np$
3. $\operatorname{Var} X = \sigma^2 = npq$

Example 3.5.2. The random variable X, the number of radar signals properly identified in a 30-minute period, is a binomial random variable with parameters $n = 10$ and $p = 1/2$. The moment generating function for this random variable is

$$m_X(t) = (1/2 + 1/2e^t)^{10}$$

Its mean is $\mu = np = 10(1/2) = 5$, and its variance is $\sigma^2 = npq = 10(1/2)(1/2) = 10/4$.

In statistical studies we shall usually be interested in computing the probability that the random variable assumes certain values. This probability can be computed from the density function, f, or from the cumulative distribution function, F. Since the binomial distribution comes into play in such a wide variety of physical applications, tables of the cumulative distribution function for selected values of n and p have been compiled. Table I of App. A is one such table. That is, Table I gives the values of

$$F(t) = \sum_{x=0}^{[t]} \binom{n}{x} p^x (1-p)^{n-x}$$

for selected values of n and p, where $[t]$ represents the greatest integer less than or equal to t. Its use is illustrated in the following example.

Example 3.5.3 Let X denote the number of radar signals properly identified in a 30-minute time period in which 10 signals are received. Assuming that X is binomial

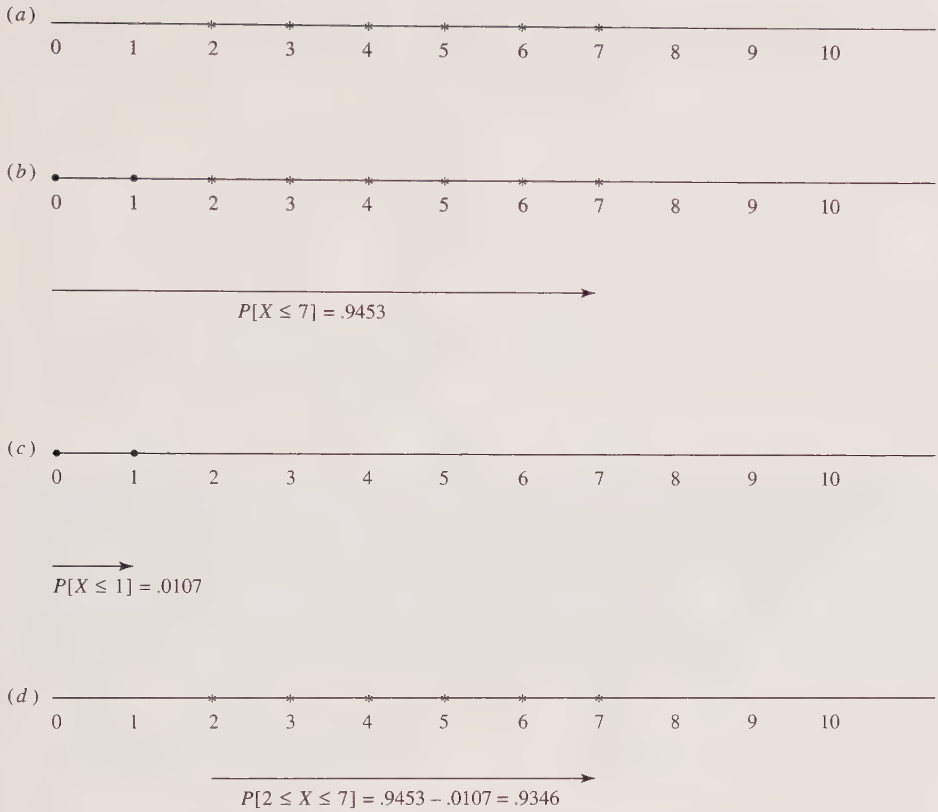

FIGURE 3.2
(*a*) The probability that X lies between 2 and 7 inclusive is the probability associated with the starred points; (*b*) $P[X \leq 7] = .9453$ includes the probability associated with 0 and 1; (*c*) the probability associated with the unwanted points 0 and 1 is .0107; (*d*) the desired probability is found by subtraction.

with $n = 10$ and $p = 1/2$, find the probability that at most seven signals will be identified correctly. This probability can be found by summing the density from $x = 0$ to $x = 7$. That is,

$$P[X \leq 7] = \sum_{x=0}^{7} \binom{10}{x} (1/2)^x (1/2)^{10-x}$$

Evaluating this probability directly entails a large amount of arithmetic. However, its value can be read from Table I of App. A. We first look at the group of values labeled $n = 10$. The desired probability of .9453 is found in the column labeled .5 and the row labeled 7. That is,

$$P[X \leq 7] = F(7) = .9453$$

Other probabilities can be found. For example, find $P[2 \leq X \leq 7]$. Figure 3.2 suggests how this is done. Notice that in Fig. 3.2 we want the probability associated with points that are starred. To determine the desired probability, we first find the number 7 in Table I of App. A. Since the table is cumulative, the probability given, .9453, is the probability that X is at most 7. This probability includes the probability that $X = 0$ or $X = 1$. Since we did not want to include those values, $P[X \leq 1] = F(1) = .0107$ must be subtracted from .9453. Thus

$$\begin{aligned} P[2 \leq X \leq 7] &= P[X \leq 7] - P[X < 2] \\ &= P[X \leq 7] - P[X \leq 1] \\ &= F(7) - F(1) \\ &= .9453 - .0107 \\ &= .9346 \end{aligned}$$

Later in the text we shall show ways of approximating binomial probabilities when the values of n and p are such that no appropriate binomial table is available.

3.6 NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution is a distribution that can be thought of as a "reversal" of the binomial distribution. In the binomial setting the random variable X represents the number of successes obtained in a series of n independent and identical Bernoulli trials; the number of trials is *fixed* and the number of successes will *vary* from experiment to experiment. The negative binomial random variable represents the number of trials needed to obtain exactly r successes; here, the number of successes is *fixed* and the number of trials will *vary* from experiment to experiment. In particular, the negative binomial random variable arises in situations characterized by the following properties:

Negative binomial properties

1. The experiment consists of a series of independent and identical Bernoulli trials, each with probability p of success.
2. The trials are observed until exactly r successes are obtained, where r is fixed by the experimenter.
3. The random variable X is the number of trials needed to obtain the r successes.

It is not hard to derive the density function for X. To do so, let us consider a setting in which $r = 3$. Typical outcomes for such an experiment are

$$ssffffs \qquad sffffss \qquad ffffsss \qquad sss \qquad ssfs$$

Here X assumes the values 7, 7, 7, 3, and 4, respectively. There are several things to notice immediately. First, each outcome must end with a successful trial. Second, the remaining $x - 1$ trials must result in exactly two successes and $x - 3$ failures *in some order*. Third, different outcomes can yield identical values for X. To determine the number of outcomes that result in a given value of X, we ask, "How many permutations can be formed consisting of $x - 1$ objects of which exactly two represent success and the rest, $x - 3$, represent failure?" The formula on page 16 can be applied to see that the answer to this question is $\binom{x-1}{2}$. For example, there are $\binom{6}{2} = 15$

ways in which X can assume the value 7. Three of these outcomes are given on page 70. Since trials are independent with probability p of success and probability $1 - p$ of failure, the probability of an outcome for which $X = x$ is given by

$$P[X = x] = \binom{x-1}{2}(1-p)^{x-3}p^3 \qquad x = 3, 4, 5, \ldots$$

You can use this expression to verify that the probability that $X = 7$ is

$$\binom{6}{2}(1-p)^4p^3$$

The argument given for $r = 3$ can be generalized easily. We simply replace 3 by r and 2 by $r - 1$ in the argument given to obtain the following definition for the negative binomial random variable:

Definition 3.6.1 (Negative binomial distribution). A random variable X is said to have a negative binomial distribution with parameters p and r if its density f is given by

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad \begin{array}{l} r = 1, 2, 3, \ldots \\ x = r, r+1, r+2, \ldots \end{array}$$

Theorem 3.6.1 gives the moment generating function for the negative binomial distribution. The expectations stated in the theorem are obtained from the moment generating function.

Theorem 3.6.1. Let X be a negative binomial random variable with parameters r and p. Then

1. the moment generating function for X is given by

$$m_X(t) = \frac{(pe^t)^r}{(1-qe^t)^r} \qquad q = 1 - p$$

2. $E[X] = r/p$
3. $\mathrm{Var}(X) = rq/p^2$

An example will illustrate the use of this distribution in a practical setting.

Example 3.6.1. Cotton linters used in the production of rocket propellant are subjected to a nitration process that enables the cotton fibers to go into solution. The process is 90% effective in that the material produced can be shaped as desired in a later processing stage with probability .9. What is the probability that exactly 20 lots will be produced in order to obtain the third defective lot? Here "success" is obtaining a defective lot, and hence $p = .1$ and $r = 3$. The probability that $X = 20$ is given by

$$f(20) = \binom{19}{2}(.9)^{17}(.1)^3$$

The expected value of X is $r/p = 3/.1$ or 30, and the variance of X is $rq/p^2 = 3(.9)/(.1)^2 = 270$. (Based on a study to compare different sources of cotton linters conducted by the Radford University Statistical Consulting Service for the Radford Army Ammunition Plant.)

One other point should be made. When $r = 1$, the negative binomial distribution reduces to the geometric distribution studied earlier. (See Exercise 51.)

3.7 HYPERGEOMETRIC DISTRIBUTION

Sampling from a finite population can be done in one of two ways. An item can be selected, examined, and returned to the population for possible reselection; or it can be selected, examined, and kept, thus preventing its reselection in subsequent draws. The former is called *sampling with replacement,* whereas the latter is called *sampling without replacement.* Sampling with replacement guarantees that the draws are independent. In sampling without replacement the draws are *not* independent. Thus if we sample without replacement, the random variable X, the number of successes in n draws, is no longer binomial. Rather, it follows a distribution known as the *hypergeometric distribution.*

Hypergeometric properties

1. The experiment consists of drawing a random sample of size n without replacement and without regard to order from a collection of N objects.
2. Of the N objects, r have a trait of interest to us; the other $N - r$ do not have the trait.
3. The random variable X is the number of objects in the sample with the trait.

To derive the density for this distribution, suppose that we have a group of N objects and that r of these objects have a trait of interest to us. We are to select n objects from the group randomly without replacement. Let X denote the number of objects chosen that have the trait. The idea is depicted in Fig. 3.3. Since we are not interested in the order in which the items are selected, we can use combinatorial techniques to conclude that there are $\binom{N}{n}$ ways to choose the n objects. In a random selection we are just as likely to obtain one set of n objects as any other. That is, there are $\binom{N}{n}$ equally likely ways in which this experiment can proceed. In order to have x successes, we must select exactly x objects from the r objects with the trait of interest. This can be done in $\binom{r}{x}$ ways. We must select the remaining $n - x$ objects from the $N - r$ objects that do not have the trait; this can be done in $\binom{N-r}{n-x}$ ways. Using classical probability and the multiplication rule for counting, we obtain

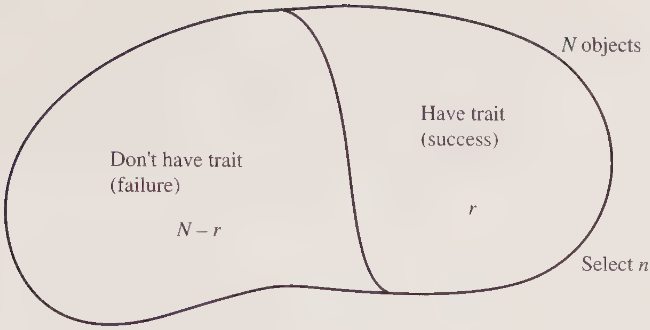

FIGURE 3.3 General hypergeometric setting.

$$P[X = x] = \frac{\text{number of ways to select } x \text{ objects with the trait and } n - x \text{ objects without the trait}}{\text{number of ways the experiment can proceed}}$$
$$= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

This argument suggests the definition of the hypergeometric distribution.

Definition 3.7.1 (Hypergeometric distribution). A random variable X has a hypergeometric distribution with parameters N, n, and r if its density is given by

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \qquad \max[0, n - (N - r)] \le x \le \min(n, r)$$

where N, r, and n are positive integers.

Notice the unusual bounds for X. A simple numerical example should show you why these bounds are as stated.

Example 3.7.1. Suppose that X is hypergeometric with $N = 15$, $r = 6$, and $n = 12$. This situation is depicted in Fig. 3.4. Since only six items have the desired trait, X cannot exceed 6. Note that $6 = \min(n, r) = \min(12, 6)$. Since we can select at most nine items from among those without the trait, we must select at least three items from among those with the trait. Note that

$$3 = \max[0, n - (N - r)] = \max[0, 12 - (15 - 6)] = \max[0, 3]$$

Just be careful when stating the bounds for a hypergeometric random variable. They are tricky! Since the bounds for X are unusual, the theoretical development of the hypergeometric distribution is not easy. However, it can be shown that

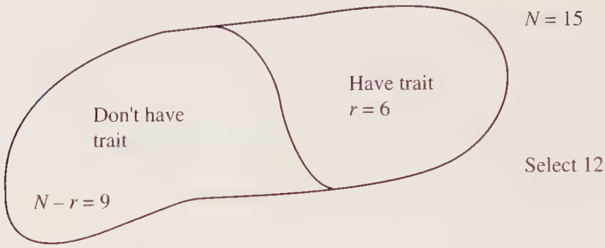

FIGURE 3.4 Hypergeometric setting with $N = 15$, $r = 6$, and $n = 12$.

$$E[X] = n\left(\frac{r}{N}\right)$$

and

$$\text{Var } X = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example 3.7.2. A foundry ships engine blocks in lots of size 20. Since no manufacturing process is perfect, defective blocks are inevitable. However, to detect the defect, the block must be destroyed. Thus we cannot test each block. Before accepting a lot, three items are selected and tested. Suppose that a given lot actually contains five defective items. Let X denote the number of defective items sampled. The density for X is

$$f(x) = \frac{\binom{5}{x}\binom{15}{3-x}}{\binom{20}{3}} \qquad x = 0, 1, 2, 3$$

The expected number of defective blocks in a sample of size 3 is

$$E[X] = n\left(\frac{r}{N}\right) = 3\left(\frac{5}{20}\right) = \frac{3}{4}$$

The variance for X is

$$\begin{aligned}\text{Var } X &= n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)\\ &= 3\left(\frac{5}{20}\right)\left(\frac{15}{20}\right)\left(\frac{17}{19}\right)\\ &= \frac{153}{304}\end{aligned}$$

If the number of items sampled (n) is small relative to the number of objects from which the sample is drawn (N), then the binomial distribution can be used to approximate hypergeometric probabilities. A rule of thumb is that the approximation is usually satisfactory if $n/N \leq .05$. The proof of this result depends upon

Stirling's formula, which is studied in courses in advanced calculus. We shall not attempt the proof here. However, the result should not be surprising. If n is small relative to N, then the composition of the sampled group does not change much from trial to trial even though we are keeping the sampled items. Thus the probability of success is not changing much from trial to trial, and for all practical purposes it can be viewed as being constant. Thus the distribution of X, the number of successes obtained in n draws, can be approximated by the binomial distribution with parameters n and $p = r/N$.

Example 3.7.3. During the course of an hour 1000 bottles of beer are filled by a particular machine. Each hour a sample of 20 bottles is randomly selected and the number of ounces of beer per bottle is checked. Let X denote the number of bottles selected that are underfilled. Suppose that during a particular hour 100 underfilled bottles are produced. Find the probability that at least 3 underfilled bottles will be among those sampled. The exact value of this probability is given by

$$\begin{aligned} P[X \geq 3] &= 1 - P[X < 3] \\ &= 1 - P[X \leq 2] \\ &= 1 - P[X=0] - P[X=1] - P[X=2] \\ &= 1 - \frac{\binom{100}{0}\binom{900}{20}}{\binom{1000}{20}} - \frac{\binom{100}{1}\binom{900}{19}}{\binom{1000}{20}} - \frac{\binom{100}{2}\binom{900}{18}}{\binom{1000}{20}} = .3224 \end{aligned}$$

As you can see, calculating this probability directly, even with the aid of a calculator, is time-consuming. However, since $n/N = 20/1000 \leq .05$, our rule of thumb indicates that this probability can be approximated by using the binomial distribution with parameters $n = 20$ and $p = r/N = 100/1000 = .1$. From Table I of App. A, the cumulative binomial table, we find that

$$\begin{aligned} P[X \geq 3] &= 1 - P[X < 3] \\ &= 1 - P[X \leq 2] \\ &= 1 - .6769 \\ &= .3231 \end{aligned}$$

3.8 POISSON DISTRIBUTION

The last discrete family to be considered is the family of *Poisson* random variables, named for the French mathematician Simeon Denis Poisson (1781–1840). The Maclaurin series expansion for the function e^z studied in beginning calculus courses provides the theoretical basis for this distribution. This series is given by

Maclaurin series

For z a real number,

$$e^z = 1 + z + z^2/2! + z^3/3! + z^4/4! + \cdots$$

We begin by considering the mathematical properties of this important family of random variables.

Definition 3.8.1 (Poisson distribution). A random variable X is said to have a Poisson distribution with parameter k if its density f is given by

$$f(x) = \frac{e^{-k}k^x}{x!} \qquad \begin{array}{l} x = 0, 1, 2, \ldots \\ k > 0 \end{array}$$

The function f given in this definition is nonnegative. To see that it sums to 1, note that

$$\sum_{x=0}^{\infty} \frac{e^{-k}k^x}{x!} = e^{-k}(1 + k + k^2/2! + k^3/3! + \cdots)$$

The series on the right is the Maclaurin series for e^k. Thus

$$\sum_{x=0}^{\infty} \frac{e^{-k}k^x}{x!} = e^{-k}e^k = e^0 = 1$$

as desired.

The moment generating function for this distribution is easy to obtain, as is its mean and variance. The following theorem gives these results. Its proof is outlined as an exercise. (Exercise 69.)

Theorem 3.8.1. Let X be a Poisson random variable with parameter k.

1. The moment generating function for X is given by

$$m_X(t) = e^{k(e^t - 1)}$$

2. $E[X] = k$
3. $\text{Var } X = k$

Poisson random variables usually arise in connection with what are called *Poisson processes*. Poisson processes involve observing discrete events in a continuous "interval" of time, length, or space. We use the word "interval" in describing the general Poisson process with the understanding that we may not be dealing with an interval in the usual mathematical sense. For example, we might observe the number of white blood cells in a drop of blood. The discrete event of interest is the observation of a white cell, whereas the continuous "interval" involved is a drop of blood. We might observe the number of times radioactive gases are emitted from a nuclear power plant during a 3-month period. The discrete event of concern is the emission of radioactive gases. The continuous interval consists of a period of 3 months. The variable of interest in a Poisson process is X, the number of occurrences of the event in an interval of length s units. Although the derivation is a bit tricky, it can be shown using differential equations that X is a Poisson random variable with parameter $k = \lambda s$, where λ is a positive number that characterizes the underlying Poisson process. To understand the physical significance of the constant λ, note that by Definition 3.8.1 the density for X is given by

$$f(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!} \qquad x = 0, 1, 2, 3, \ldots$$

By Theorem 3.8.1 the expected value of X is λs. That is, the average number of occurrences of the event of interest in an interval of s units is λs. Thus the average number of occurrences of the event in 1 unit of time, length, area, or space is $\lambda s/s = \lambda$. That is, physically, the *parameter λ of a Poisson process represents the average number of occurrences of the event in question per measurement unit.*

The following steps are used in the solution of an applied Poisson problem:

Steps in Solving a Poisson Problem

1. Determine the basic unit of measurement being used.
2. Determine the average number of occurrences of the event per unit. This number is denoted by λ.
3. Determine the length or size of the observation period. This number is denoted by s.
4. The random variable X, the number of occurrences of the event in the interval of size s follows a Poisson distribution with parameter $k = \lambda s$.

These steps are illustrated in Example 3.8.1.

Example 3.8.1. The white blood cell count of a healthy individual can average as low as 6000 per cubic millimeter of blood. To detect a white-cell deficiency, a .001 cubic millimeter drop of blood is taken and the number of white cells X is found. How many white cells are expected in a healthy individual? If at most two are found, is there evidence of a white cell deficiency?

This experiment can be viewed as involving a Poisson process. The discrete event of interest is the occurrence of a white cell; the continuous interval is a drop of blood.

Let the measurement unit be a cubic millimeter; then $s = .001$ and λ, the average number of occurrences of the event per unit, is 6000. Thus X is a Poisson random variable with parameter $\lambda s = 6000(.001) = 6$. By Theorem 3.8.1, $E[X] = \lambda s = 6$. In a healthy individual we would expect, on the average, to see six white cells. How rare is it to see at most two? That is, what is $P[X \leq 2]$? From Definition 3.8.1,

$$P[X \leq 2] = \sum_{x=0}^{2} f(x) = \sum_{x=0}^{2} \frac{e^{-6}6^x}{x!}$$

$$= \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!}$$

Evaluating this type of expression directly does entail some arithmetic.

Once again, because of the wide appeal of the Poisson model, the values of the cumulative distribution function for selected values of the parameter $k = \lambda s$ are tabulated. Table II of App. A is one such table. The desired probability of .062 is found by

TABLE 3.8
Discrete distributions: A summary

Name	Density		Moment generating function	Mean	Variance
Geometric	$(1-p)^{x-1}p$	$x = 1, 2, 3, \ldots$ $0 < p < 1$	$\dfrac{pe^t}{1-qe^t}$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$
Uniform	$\dfrac{1}{n}$	$x = x_1, x_2, \ldots, x_n$ n a positive integer	$\dfrac{\sum_{i=1}^{n} e^{tx_i}}{n}$	$\dfrac{\sum_{i=1}^{n} x_i}{n}$	$\dfrac{\sum_{i=1}^{n} x_i^2}{n} - \left(\dfrac{\sum_{i=1}^{n} x_i}{n}\right)^2$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$x = 0, 1, 2, \ldots, n$ $0 < p < 1$ n a positive integer	$(q + pe^t)^n$	np	$np(1-p)$
Bernoulli or point binomial	$p^x(1-p)^{1-x}$	$x = 0, 1$ $0 < p < 1$	$q + pe^t$	p	$p(1-p)$
Hypergeometric	$\dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$\max[0, n-(N-r)]$ $\leq x \leq \min(n, r)$		$n\dfrac{r}{N}$	$n\dfrac{r}{N}\left(\dfrac{N-r}{N}\right)\left(\dfrac{N-n}{N-1}\right)$
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$x = r, r+1, r+2, \ldots,$ $0 < p < 1$	$\dfrac{(pe^t)^r}{(1-qe^t)^r}$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$
Poisson	$\dfrac{e^{-k}k^x}{x!}$	$x = 0, 1, 2, \ldots$ $k > 0$	$e^{k(e^t-1)}$	k	k

looking under the column labeled $k = 6$ in the row labeled 2. Is there evidence of a white-cell deficiency? There are no rules that say at what point probabilities are considered to be small. To answer this question, a value judgment must be made. If you consider .062 to be small, then the natural conclusion is that the individual does have a white-cell deficiency.

3.9 SIMULATING A DISCRETE DISTRIBUTION

In designing operating systems of various types, one often needs to simulate the system before it is built. Simulation is usually done with the aid of a computer. However, the idea behind simulation can be illustrated by using a random digit table. A portion of such a table is given in Table III of App. A. Its use is illustrated in the following example.

Example 3.9.1. Table 3.9 presents a portion of the random digit table in the appendix. Let us read a sequence of random two-digit numbers from this table. To do so, we must get a random start. This can be done by writing the integers 1 through 14 on slips of paper, placing the slips in a bowl, stirring, and drawing one slip at random from the bowl. The number selected identifies the column in which our starting number is located. In a similar way, we can select the row in which the starting number is located.

TABLE 3.9

Row	Column / Random digits (1)	(2)	(3)
1	10480	15011	01536
2	22368	46573	25595
3	24130	48360	22527
4	42167	93093	06243
5	37570	39975	81837
6	77921	06907	11008
7	99562	72905	56420
8	96301	91977	05463
9	89579	14342	63661
10	85485	36857	43342

Suppose that this process results in the selection of column 2 and row 5. This identifies the random starting point as 39975.

Since we want two-digit numbers, we need only read the first two digits of this number. Thus our first random number is 39. Since a random digit table is constructed in such a way that the digit appearing at each position in the table is just as likely to be one digit as any other, the table can be read in any way. Let us agree to read down the second column so that the next four two-digit numbers are 06, 72, 91, and 14.

The next example illustrates the use of a random digit table in a simple simulation experiment.

Example 3.9.2. Suppose that at a particular airport planes arrive at an average rate of one per minute and depart at the same average rate. We are interested in simulating the behavior of the random variable Z, the number of planes on the ground at a given time. We will simulate Z for five consecutive one-minute periods. Note that for each of these periods the random variables X, the number of arrivals, and Y, the number of departures, are both Poisson variables with parameter $k = 1$. The density for X and Y is obtained from Table II of App. A and is shown below:

$$P[X = 0] = P[Y = 0] = .368$$
$$P[X = 1] = P[Y = 1] = .368$$
$$P[X = 2] = P[Y = 2] = .184$$
$$P[X = 3] = P[Y = 3] = .061$$
$$P[X = 4] = P[Y = 4] = .015$$
$$P[X = 5] = P[Y = 5] = .003$$
$$P[X = 6] = P[Y = 6] = .001$$
$$P[X > 6] = P[Y > 6] \doteq 0$$

There are 1000 possible three-digit numbers. We divide them into seven categories to reflect the above probabilities. This division is shown in Table 3.10. To perform the simulation, we read a total of 10 random three-digit numbers using the procedure demonstrated in Example 3.9.1. Assume that at the beginning of the simulation there

TABLE 3.10

Random number	Number of arrivals (x)	Number of departures (y)	$P[X = x] = P[Y = y]$
000–367	0	0	.368
368–735	1	1	.368
736–919	2	2	.184
920–980	3	3	.061
981–995	4	4	.015
996–998	5	5	.003
999	6	6	.001

TABLE 3.11

Time span, min	Random 3-digit number	Number of arrivals (x)	Number of departures (y)	Number on ground at end of time period (z)
1	015	0		100
	255		0	100
2	225	0		
	062		0	100
3	818	2		
	110		0	102
4	564	1		
	054		0	103
5	636	1		
	433		1	103

are 100 planes on the ground and that our random starting point is the number <u>015</u>36 found in line 1 and column 3 of Table 3.9. The first number read corresponds to the arrivals during the first minute of observation, the second to the departures during this time span, and so forth. The results of the simulation are shown in Table 3.11. If this simulation were continued over a long period of time, we could begin to answer such questions as: "On the average, how many planes are on the ground at a given time?" and "How much variability is there in the number of planes on the ground?"

CHAPTER SUMMARY

In this chapter we introduced the concept of a random variable and showed you how to distinguish a discrete random variable from one that is not discrete. We studied two functions, the density function and the cumulative distribution function, that are used to compute probabilities. The density gives the probability that X assumes a specific value x; the cumulative distribution gives the probability that X assumes a value less than or equal to x. The concept of expected value was introduced and used to define three important parameters, the mean (μ), the variance (σ^2), and the standard deviation (σ). The mean is a measure of the center of location of the distribution; the variance and standard deviation measure the variability of the random variable about its mean. The moment generating function was introduced as a means of finding the mean and variance of X. Special discrete distributions that find extensive use in all areas of application were presented. These are the geometric, hypergeometric, negative binomial, binomial, Bernoulli, uniform, and Poisson distributions. We also discussed briefly how to simulate a discrete distribution. We introduced and defined terms that you should know. These are:

Random variable
Discrete random variable
Discrete density
Cumulative distribution
Expected value
Mean
Variance
Standard deviation
Bernoulli trial
Moment generating function
Sampling with replacement
Sampling without replacement

EXERCISES

Section 3.1

In each of the following, identify the variable as discrete or not discrete.

1. *T*: the turnaround time for a computer job (the time it takes to run the program and receive the results).
2. *M*: the number of meteorites hitting a satellite per day.
3. *N*: the number of neutrons expelled per thermal neutron absorbed in fission of uranium-235.
4. Neutrons emitted as a result of fission are either prompt neutrons or delayed neutrons. Prompt neutrons account for about 99% of all neutrons emitted and are released within 10^{-14} s of the instant of fission. Delayed neutrons are emitted over a period of several hours. Let D denote the time at which a delayed neutron is emitted in a fission reaction.
5. Electrical resistance is the opposition offered by electrical conductors to the flow of current. The unit of resistance is the ohm. For example, a 2½-inch electric bell will usually have a resistance somewhere between 1.5 and 3 ohms. Let O denote the actual resistance of a randomly selected bell of this type.
6. The number of power failures per month in the Tennessee Valley power network.

Section 3.2

7. Grafting, the uniting of the stem of one plant with the stem or root of another, is widely used commercially to grow the stem of one variety that produces fine fruit on the root system of another variety with a hardy root system. Most Florida sweet oranges grow on trees grafted to the root of a sour orange variety. The density for X, the number of grafts that fail in a series of five trials, is given by Table 3.12.

TABLE 3.12

x	0	1	2	3	4	5
$f(x)$	.7	.2	.05	.03	.01	?

TABLE 3.13

x	1	2	3	4	5	6	7	8
$f(x)$	.02	.03	.05	.2	.4	.2	.07	?

(*a*) Find $f(5)$.
(*b*) Find the table for F.
(*c*) Use F to find the probability that at most three grafts fail; that at least two grafts fail.
(*d*) Use F to verify that the probability of exactly three failures is .03.

8. In blasting soft rock such as limestone, the holes bored to hold the explosives are drilled with a Kelly bar. This drill is designed so that the explosives can be packed into the hole before the drill is removed. This is necessary since in soft rock the hole often collapses as the drill is removed. The bits for these drills must be changed fairly often. Let X denote the number of holes that can be drilled per bit. The density for X is given in Table 3.13.
(*a*) Find $f(8)$.
(*b*) Find the table for F.
(*c*) Use F to find the probability that a randomly selected bit can be used to drill between three and five holes inclusive.
(*d*) Find $P[X \leq 4]$ and $P[X < 4]$. Are these probabilities the same?
(*e*) Find $F(-3)$ and $F(10)$. *Hint:* Express these in terms of the probabilities that they represent and their values will become obvious.

9. Consider Example 1.2.1. Let X denote the number of computer systems operable at the time of the launch. Assume that the probability that each system is operable is .9.
(*a*) Use the tree of Fig. 1.2 to find the density table.
(*b*) There is a pattern to the probabilities in the density table. In particular,

$$f(x) = k(x)(.9)^x(.1)^{3-x}$$

where $k(x)$ gives the number of paths through the tree yielding a particular value for X. Verify that $k(x) = \binom{n}{x}$ for $x = 0, 1, 2, 3$
(*c*) Find the table for F.
(*d*) Use F to find the probability that at least one system is operable at launch time.
(*e*) Use F to find the probability that at most one system is operable at the time of the launch.

10. It is known that the probability of being able to log on to a computer from a remote terminal at any given time is .7. Let X denote the number of attempts that must be made to gain access to the computer.
(*a*) Find the first four terms of the density table.
(*b*) Find a closed-form expression for $f(x)$.
(*c*) Find $P[X = 6]$.
(*d*) Find a closed-form expression for $F(x)$.
(*e*) Use F to find the probability that at most four attempts must be made to gain access to the computer.

TABLE 3.14

x	0	1	2	3	4	5	6
$F(x)$	.05	.15	.35	.65	.85	.95	1.0

(*f*) Use F to find the probability that at least five attempts must be made to gain access to the computer.

11. Knitting machines at a factory making elastic use a laser to detect broken threads. When a thread breaks, the machine must be stopped and the broken thread must be found and repaired by a technician. Assume that the density for X, the number of times per day that a specific machine is stopped, is given by

$$f(x) = \left(\frac{16}{31}\right)\left(\frac{1}{2}\right)^x \qquad x = 0, 1, 2, 3, 4$$

(*a*) Find the density table for X, and verify that the sum of the probabilities given in the table is 1.
(*b*) If $x < 0$, what is the numerical value of $F(x)$?
(*c*) If $x > 4$, what is the numerical value of $F(x)$?

12. Past experience shows that over time the rivets in bridge supports can become dangerously loose. Assume that X, the number of loose rivets found per 10 feet beam on bridges over 20 years old, has the cumulative distribution shown in Table 3.14.
(*a*) Find the density table for X.
(*b*) Verify that $f(x) = \dfrac{6 - 2|x - 3|}{20} \qquad x = 1,2,3,4,5$

$$f(x) = \frac{4 - |x - 3|}{20} \qquad x = 0 \text{ or } 6$$

13. Explain why the cumulative distribution function for a discrete random variable can never decrease in value.

Section 3.3

14. In an experiment to graft Florida sweet orange trees to the root of a sour orange variety, a series of five trials is conducted. Let X denote the number of grafts that fail. The density for X is given in Table 3.12.
(*a*) Find $E[X]$.
(*b*) Find μ_X.
(*c*) Find $E[X^2]$.
(*d*) Find Var X.
(*e*) Find σ_X^2.
(*f*) Find the standard deviation for X.
(*g*) What physical unit is associated with σ_X?

15. The density for X, the number of holes that can be drilled per bit while drilling into limestone is given in Table 3.13.
(*a*) Find $E[X]$ and $E[X^2]$.
(*b*) Find Var X and σ_X.
(*c*) What physical unit is associated with σ_X?

16. Use the density derived in Exercise 9 to find the expected value and variance for X, the number of computer systems operable at the time of the launch. Can you express $E[X]$ and $\operatorname{Var} X$ in terms of n, the number of systems available, and p, the probability that a given system will be operable?

17. The probability p of being able to log on to a computer from a remote terminal at any given time is .7. Let X denote the number of attempts that must be made to gain access to the computer. Find $E[X]$. Can you express $E[X]$ in terms of p? *Hint:* The series $\sum_{x=1}^{\infty} x(.7)(.3)^{x-1} = E[X]$ is not geometric. To find $E[X]$, expand this series and the series $.3E[X]$. Subtract the two to form the series $.7E[X]$. Evaluate this *geometric* series, and solve for $E[X]$.

18. The probability that a cell will fuse in the presence of polyethylene glycol is 1/2. Let Y denote the number of cells exposed to antigen-carrying lymphocytes to obtain the first fusion. Use the method of Exercise 17 to find $E[Y]$.

19. Let X be a discrete random variable with density f. Let c be any real number. Show that

(*a*) $E[c] = c$. *Hint:* Remember that constants can be factored from summations and that $\sum_{\text{all } x} f(x) = 1$.

(*b*) $E[cX] = cE[X]$.

20. Use the rules for expectation to verify that $\operatorname{Var} c = 0$ and $\operatorname{Var} cX = c^2 \operatorname{Var} X$ for any real number c. *Hint:* $\operatorname{Var} c = E[c^2] - (E[c])^2$.

21. Let X and Y be independent random variables with $E[X] = 3$, $E[X^2] = 25$, $E[Y] = 10$ and $E[Y^2] = 164$.

(*a*) Find $E[3X + Y - 8]$.

(*b*) Find $E[2X - 3Y + 7]$.

(*c*) Find $\operatorname{Var} X$.

(*d*) Find σ_X.

(*e*) Find $\operatorname{Var} Y$.

(*f*) Find σ_Y.

(*g*) Find $\operatorname{Var}[3X + Y - 8]$.

(*h*) Find $\operatorname{Var}[2X - 3Y + 7]$.

(*i*) Find $E[(X - 3)/4]$ and $\operatorname{Var}[(X - 3)/4]$.

(*j*) Find $E[(Y - 10)/8]$ and $\operatorname{Var}[(Y - 10)/8]$.

(*k*) The results of parts (*i*) and (*j*) are not coincidental. Can you generalize and verify the conjecture suggested by these two exercises?

22. Consider the function f defined by

$$f(x) = (1/2)2^{-|x|} \qquad x = \pm 1, \pm 2, \pm 3, \pm 4, \ldots$$

(*a*) Verify that this is the density for a discrete random variable X. *Hint:* Expand the series $\sum_{\text{all } x} f(x)$ for a few terms. A recognizable series will develop!

(*b*) Let $g(X) = (-1)^{|X|-1}[2^{|X|}/(2|X| - 1)]$. Show that $\sum_{\text{all } x} g(x)f(x) < \infty$. *Hint:* Expand the series for a few terms. You will obtain an alternating series that can be shown to converge.

(*c*) Show that $\sum_{\text{all } x} |g(x)| f(x)$ does not converge. This will show that $E[g(X)]$ does not exist. *Hint:* Expand the series for a few terms. You will obtain a series that is term by term larger than the diverging harmonic type series $(1/3)\sum_{x=1}^{\infty} 1/x$.

23. *(An application to sort algorithms.)* In studying various sort algorithms in computer science, it is of interest to compare their efficiency by estimating the average number of interchanges needed to sort random arrays of various sizes. It is also of interest to compare these estimated averages to the "ideal" average, where by "ideal" we mean the expected minimum number of interchanges needed to sort the array. In this exercise you will derive this ideal average. (American Mathematical Association of Two-Year Colleges, "A Note on the Minimum Number of Interchanges Needed to Sort a Random Array," with T. McMillan, I. Liss, and J. Milton, Fall 1990.)

(*a*) Consider a random array of length n. When the positions of exactly two elements of the array are exchanged, we say that an "interchange" has taken place. Let X_n denote the minimum number of interchanges necessary to sort an array of size n. Note that

$$X_n = X_{n-1} + I$$

where $I = 0$ if the last element of the array is in the correct position and $I = 1$ otherwise. Argue that $P[I = 0] = 1/n$ and $P[I = 1] = 1 - (1/n)$.

(*b*) Show that

$$E[I] = 1 - \frac{1}{n}$$

(*c*) Argue that

$$\begin{aligned}
E[X_n] &= E[X_{n-1}] + 1 - \frac{1}{n} \\
E[X_{n-1}] &= E[X_{n-2}] + 1 - \frac{1}{n-1} \\
E[X_{n-2}] &= E[X_{n-3}] + 1 - \frac{1}{n-2} \\
&\vdots \\
E[X_3] &= E[X_2] + 1 - \frac{1}{3} \\
E[X_2] &= E[X_1] + 1 - \frac{1}{2} \\
E[X_1] &= 0
\end{aligned}$$

(*d*) Use a recursive argument to show that

$$E[X_n] = (n-1) - \sum_{i=2}^{n} \frac{1}{i}$$

(*e*) Illustrate the expression given in part (*d*) by finding $E[X_5]$.

(*f*) Elementary calculus can be used to approximate $E[X_n]$ by noting that

$$\sum_{i=2}^{n} \frac{1}{i} \doteq \int_{1.5}^{n+.5} \frac{1}{t}\, dt$$

Use this idea to approximate $E[X_5]$ and to compare the result to the exact solution found in part (*e*).

(*g*) A random digit generator is used to generate sets of 100 different three-digit numbers lying between 0 and 1. What is the ideal average number of interchanges needed to sort such an array?

Section 3.4

24. The probability that a wildcat well will be productive is 1/13. Assume that a group is drilling wells in various parts of the country so that the status of one well has no bearing on that of any other. Let X denote the number of wells drilled to obtain the first strike.

(*a*) Verify that X is geometric, and identify the value of the parameter p.

(*b*) What is the exact expression for the density for X?

(*c*) What is the exact expression for the moment generating function for X?

(*d*) What are the numerical values of $E[X]$, $E[X^2]$, σ^2, and σ?

(e) Find $P[X \geq 2]$.

25. The zinc-phosphate coating on the threads of steel tubes used in oil and gas wells is critical to their performance. To monitor the coating process, an uncoated metal sample with known outside area is weighed and treated along with the lot of tubing. This sample is then stripped and reweighed. From this it is possible to determine whether or not the proper amount of coating was applied to the tubing. Assume that the probability that a given lot is unacceptable is .05. Let X denote the number of runs conducted to produce an unacceptable lot. Assume that the runs are independent in the sense that the outcome of one run has no effect on that of any other.

(*a*) Verify that X is geometric. What is "success" in this experiment? What is the numerical value of p?

(*b*) What is the exact expression for the density for X?

(*c*) What is the exact expression for the moment generating function for X?

(*d*) What are the numerical values of $E[X]$, $E[X^2]$, σ^2, and σ?

(*e*) Find the probability that the number of runs required to produce an unacceptable lot is at least 3.

26. Let X be geometric with probability of success p. Prove that when x is a positive integer, $F(x) = 1 - q^x$. Verify that this result holds true for the density given in Example 3.2.4. Argue that, in general, $F(x) = 1 - q^{[x]}$.

27. Find the expression for the cumulative distribution function for the random variable of Exercise 25. Use this function to find the probability that at most three runs are required to produce an unacceptable lot.

28. A system used to read electric meters automatically requires the use of a 128-bit computer message. Occasionally random interference causes a digit reversal resulting in a transmission error. Assume that the probability of a digit reversal for each bit is 1/1000. Let X denote the number of transmission errors per 128-bit message sent. Is X geometric? If not, what geometric property fails?

29. Verify that the random variable X of Exercise 17 is geometric. Use Theorem 3.4.3 to find $E[X]$, and compare your answer to that obtained in Exercise 17.

30. Verify that the random variable Y of Exercise 18 is geometric. Use Theorem 3.4.3 to find $E[Y]$, and compare your answer to that obtained in Exercise 18.

31. Consider the random variable X whose density is given by

$$f(x) = \frac{(x-3)^2}{5} \qquad x = 3, 4, 5$$

(*a*) Verify that this function is a density for a discrete random variable.
(*b*) Find $E[X]$ directly. That is, evaluate $\sum_{\text{all } x} xf(x)$.
(*c*) Find the moment generating function for X.
(*d*) Use the moment generating function to find $E[X]$, thus verifying your answer to part (*b*) of this exercise.
(*e*) Find $E[X^2]$ directly. That is, evaluate $\sum_{\text{all } x} x^2 f(x)$.
(*f*) Use the moment generating function to find $E[X^2]$, thus verifying your answer to part (*e*) of this exercise.
(*g*) Find σ^2 and σ.

32. A discrete random variable has moment generating function

$$m_X(t) = e^{2(e^t - 1)}$$

(*a*) Find $E[X]$.
(*b*) Find $E[X^2]$.
(*c*) Find σ^2 and σ.

33. A quality engineer is monitoring a process that produces timing belts for automobiles. Each hour he samples 4 belts from the production line and determines the average breaking strength for the sample. If the average is too low, then this is a signal that the process is not operating correctly and that adjustments need to be made. Assume that when the process is working correctly the probability of obtaining a sample that produces an average that is too low is .025. Assume that this probability remains the same for each sample drawn.
(*a*) Argue that X, the number of samples that are drawn in order to obtain the first sample that produces an average that is too low, follows the geometric distribution, and identify the numerical value of p.
(*b*) Write the formula for the moment generating function for X.
(*c*) On the average, how many samples will be drawn in order to obtain the first sample whose average is too low?

34. *(Discrete uniform distribution.)* A discrete random variable is said to be *uniformly* distributed if it assumes a finite number of values with each value occurring with the same probability. If we consider the generation of a single random digit, then Y, the number generated, is uniformly distributed with each possible digit occurring with probability 1/10. In general, the density for a uniformly distributed random variable is given by

$$f(x) = 1/n \qquad \begin{array}{l} n \text{ a positive integer} \\ x = x_1, x_2, x_3, \ldots, x_n \end{array}$$

(*a*) Find the moment generating function for a discrete uniform random variable.
(*b*) Use the moment generating function to find $E[X]$, $E[X^2]$, and σ^2.

(*c*) Find the mean and variance for the random variable Y, the number obtained when a random digit generator is activated once. *Hint:* The sum of the first n positive integers is $n(n+1)/2$; the sum of the squares of the first n positive integers is $n(n+1)(2n+1)/6$.

35. Let the density for X be given by

$$f(x) = ce^{-x} \qquad x = 1, 2, 3, \ldots$$

(*a*) Find the value of c that makes this a density.
(*b*) Find the moment generating function for X.
(*c*) Use $m_X(t)$ to find $E[X]$.

Section 3.5

36. Let X be binomial with parameters $n = 15$ and $p = .2$.
(*a*) Find the expression for the density for X.
(*b*) Find the expression for the moment generating function for X.
(*c*) Find $E[X]$ and Var X.
(*d*) Find $E[X]$, $E[X^2]$, and Var X using the moment generating function, thus verifying your answer to part (*c*) of this exercise.
(*e*) Find $P[X \leq 1]$ by evaluating the density directly. Compare your answer to that given in Table I of App. A.
(*f*) Draw dot diagrams similar to that of Fig. 3.2 to illustrate each of these probabilities, and find the probabilities using Table I of App. A.

$$\begin{array}{ll} P[X \leq 5] & P[X \geq 3] \\ P[X < 5] & F(9) \\ P[2 \leq X \leq 7] & F(20) \\ P[2 \leq X < 7] & P[X = 10] \end{array}$$

37. Albino rats used to study the hormonal regulation of a metabolic pathway are injected with a drug that inhibits body synthesis of protein. The probability that a rat will die from the drug before the experiment is over is .2. If 10 animals are treated with the drug, how many are expected to die before the experiment ends? What is the probability that at least eight will survive? Would you be surprised if at least five died during the course of the experiment? Explain, based on the probability of this occurring.

38. Consider Example 1.2.1. The random variable X is the number of computer systems operable at the time of a space launch. The systems are assumed to operate independently. Each is operable with probability .9.
(*a*) Argue that X is binomial and find its density. Compare your answer to that obtained in Exercise 9(*b*).
(*b*) Find $E[X]$ and Var X.

39. In humans, geneticists have identified two sex chromosomes, R and Y. Every individual has an R chromosome, and the presence of a Y chromosome distinguishes the individual as male. Thus the two sexes are characterized as RR (female) and RY (male). Color blindness is caused by a recessive allele on the R chromosome, which we denote by r. The Y chromosome has no bearing on color blindness. Thus relative to color blindness, there are three genotypes for females and two for males:

Female	Male
RR (normal)	*RY* (normal)
Rr (carrier)	*rY* (color-blind)
rr (color-blind)	

A child inherits one sex chromosome randomly from each parent.

(*a*) A carrier of color blindness parents a child with a normal male. Construct a tree to represent the possible genotypes for the child. Use the tree to find the probability that a given child will be a color-blind male.

(*b*) If the couple has five children, what is the expected number of color-blind males? What is the probability that three or more will be color-blind males?

40. In scanning electron microscopy photography, a specimen is placed in a vacuum chamber and scanned by an electron beam. Secondary electrons emitted from the specimen are collected by a detector, and an image is displayed on a cathode-ray tube. This image is photographed. In the past a 4- × 5-inch camera has been used. It is thought that a 35-millimeter (mm) camera can obtain the same clarity. This type of camera is faster and more economical than the 4- × 5-inch variety.

(*a*) Photographs of 15 specimens are made using each camera system. These unmarked photographs are judged for clarity by an impartial judge. The judge is asked to select the better of the two photographs from each pair. Let X denote the number selected taken by a 35-mm camera. If there is really no difference in clarity and the judge is randomly selecting photographs, what is the expected value of X?

(*b*) Would you be surprised if the judge selected 12 or more photographs taken by the 35-mm camera? Explain, based on the probability involved.

(*c*) If $X \geq 12$, do you think that there is reason to suspect that the judge is not selecting the photographs at random?

41. It has been found that 80% of all printers used on home computers operate correctly at the time of installation. The rest require some adjustment. A particular dealer sells 10 units during a given month.

(*a*) Find the probability that at least nine of the printers operate correctly upon installation.

(*b*) Consider 5 months in which 10 units are sold per month. What is the probability that at least 9 units operate correctly in each of the 5 months?

42. It is possible for a computer to pick up an erroneous signal that does not show up as an error on the screen. The error is called a silent paging error. A particular terminal is defective, and when using the system word processor, it introduces a silent paging error with probability .1. The word processor is used 20 times during a given week.

(*a*) Find the probability that no silent paging errors occur.

(*b*) Find the probability that at least one such error occurs.
(*c*) Would it be unusual for more than four such errors to occur? Explain, based on the probability involved.

43. (*a*) Find the moment generating function for a binomial random variable with parameters n and p. *Hint*: Let

$$\binom{n}{x}e^{tx}p^x(1-p)^{n-x} = \binom{n}{x}(pe^t)^x(1-p)^{n-x}$$

and apply the binomial theorem.
(*b*) Use $m_X(t)$ to show that $E[X] = np$.
(*c*) Use $m_X(t)$ to show that $E[X^2] = n^2p^2 - np^2 + np$.
(*d*) Show that Var $X = npq$, where $q = 1 - p$.

44. Assume that each time a metal detector at an airport signals, there is a 25% chance that the cause is change in the passenger's pocket. During a given hour, 15 passengers are stopped because of a signal from the metal detector.
(*a*) Find the probability that at least 3 persons will have been stopped due to change in their pockets.
(*b*) If 15 passengers are stopped by the detector, would it be unusual for none of these to have been stopped due to change in the pocket? Explain based on the probability of this occurring.

45. *(Point binomial or Bernoulli distribution.)* Assume that an experiment is conducted and that the outcome is considered to be either a success or a failure. Let p denote the probability of success. Define X to be 1 if the experiment is a success and 0 if it is a failure. X is said to have a *point binomial* or a *Bernoulli* distribution with parameter p.
(*a*) Argue that X is a binomial random variable with $n = 1$.
(*b*) Find the density for X.
(*c*) Find the moment generating function for X.
(*d*) Find the mean and variance for X.
(*e*) In DNA replication errors can occur that are chemically induced. Some of these errors are "silent" in that they do not lead to an observable mutation. Growing bacteria are exposed to a chemical that has probability .14 of inducing an observable error. Let X be 1 if an observable mutation results, and let X be 0 otherwise. Find $E[X]$.

46. A binomial random variable has mean 5 and variance 4. Find the values of n and p that characterize the distribution of this random variable.

Section 3.6

47. A company is manufacturing highway emergency flares. Such flares are supposed to burn for an average of 20 minutes. Every hour a sample of flares is collected, and their average burn time is determined. If the manufacturing process is working correctly, there is a 68% chance that the average burn time of the sample will be between 14 minutes and 26 minutes. The quality engineer in charge of the process believes that if 4 of 5 samples fall outside these bounds then this is a signal that the process might not be performing as expected. Each morning the sampling begins anew. Let X denote the number of samples drawn in order to obtain the fourth sample whose average value is outside of the above bounds. Find the probability that for a given morning $X = 5$ and hence there seems to be a problem right away.

48. A particular pitching machine is manufactured so that it will throw the ball into the strike zone of a 6-foot batter 90% of the time. What is the average number of pitches that it will throw in order to walk a batter (that is, throw 4 pitches outside of the strike zone)? What is the probability that the fourth ball will be thrown on the seventh pitch?

49. Use the moment generating function to show that the mean of a negative binomial distribution with parameters r and p is r/p.

50. Use the moment generating function to show that $E[X^2] = (r^2 + rq)/p^2$ and that $\text{Var } X = rq/p^2$ for the negative binomial distribution with parameters r and p.

51. Show that the geometric distribution is a special case of the negative binomial distribution with $r = 1$. Find the mean and variance of a geometric random variable with parameter p using Exercises 49 and 50. Compare your answer with the results of Theorem 3.4.3.

52. A vaccine for desensitizing patients to bee stings is to be packed with three vials in each box. Each vial is checked for strength before packing. The probability that a vial meets specifications is .9. Let X denote the number of vials that must be checked to fill a box. Find the density for X and its mean and variance. Would you be surprised if seven or more vials have to be tested to find three that meet specifications? Explain, based on the probability of this occurrence.

53. Some characteristics in animals are said to be sex-influenced. For example, the production of horns in sheep is governed by a pair of alleles, H and h. The allele H for the production of horns is dominant in males but recessive in females. The allele h for hornlessness is dominant in females and recessive in males. Thus, given a heterozygous male (Hh) and a heterozygous female (Hh), the male will have horns but the female will be hornless. Assume that two such animals mate and the offspring is just as likely to be male as female. The lamb inherits one gene for horns randomly from each parent. Use a tree diagram to show that the probability that a lamb will be a hornless female is 3/8. Find the average number of lambs born to obtain the second hornless female. Would you be surprised if at most five lambs were born to obtain the second hornless female? Explain.

Section 3.7

54. Suppose that X is hypergeometric with $N = 20$, $r = 17$, and $n = 5$. What are the possible values for X? What is $E[X]$ and $\text{Var } X$?

55. Suppose that X is hypergeometric with $N = 20$, $r = 3$, and $n = 5$. What are the possible values for X? What is $E[X]$ and $\text{Var } X$?

56. Suppose that X is hypergeometric with $N = 20$, $r = 10$, and $n = 5$. What are the possible values for X? What is $E[X]$ and $\text{Var } X$?

57. Twenty microprocessor chips are in stock. Three have etching errors that cannot be detected by the naked eye. Five chips are selected and installed in field equipment.

(*a*) Find the density for X, the number of chips selected that have etching errors.

(*b*) Find $E[X]$ and Var X.
(*c*) Find the probability that no chips with etching errors will be selected.
(*d*) Find the probability that at least one chip with an etching error will be chosen.

58. Production line workers assemble 15 automobiles per hour. During a given hour, four are produced with improperly fitted doors. Three automobiles are selected at random and inspected. Let X denote the number inspected that have improperly fitted doors.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Find the probability that at most one will be found with improperly fitted doors.

59. A distributor of computer software wants to obtain some customer feedback concerning its newest package. Three thousand customers have purchased the package. Assume that 600 of these customers are dissatisfied with the product. Twenty customers are randomly sampled and questioned about the package. Let X denote the number of dissatisfied customers sampled.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Set up the calculations needed to find $P[X \leq 3]$.
(*d*) Use the binomial tables to approximate $P[X \leq 3]$.

60. A random telephone poll is conducted to ascertain public opinion concerning the construction of a nuclear power plant in a particular community. Assume that there are 150,000 numbers listed for private individuals and that 90,000 of these would elicit a negative response if contacted. Let X denote the number of negative responses obtained in 15 calls.
(*a*) Find the density for X.
(*b*) Find $E[X]$ and Var X.
(*c*) Set up the calculations needed to find $P[X \geq 6]$.
(*d*) Use the binomial tables to approximate $P[X \geq 6]$.

Section 3.8

61. Let X be a Poisson random variable with parameter $k = 10$.
(*a*) Find $E[X]$.
(*b*) Find Var X.
(*c*) Find σ_X.
(*d*) Find the expression for the density for X.
(*e*) Find $P[X \leq 4]$.
(*f*) Find $P[X < 4]$.
(*g*) Find $P[X = 4]$.
(*h*) Find $P[X \geq 4]$.
(*i*) Find $P[4 \leq X \leq 9]$.

62. A particular nuclear plant releases a detectable amount of radioactive gases twice a month on the average. Find the probability that there will be at most four such emissions during a month. What is the expected number of emissions during a 3-month period? If, in fact, 12 or more emissions are detected during a 3-month period, do you think that there is a reason to suspect the reported average figure of twice a month? Explain, on the basis of the probability involved.

63. Geophysicists determine the age of a zircon by counting the number of uranium fission tracks on a polished surface. A particular zircon is of such an age that the average number of tracks per square centimeter is five. What is the probability that a 2-centimeter-square sample of this zircon will reveal at most three tracks, thus leading to an underestimation of the age of the material?

64. California is hit by approximately 500 earthquakes that are large enough to be felt every year. However, those of destructive magnitude occur on the average once every year. Find the probability that California will experience at least one earthquake of this magnitude during a 6-month period. Would it be unusual to have 3 or more earthquakes of destructive magnitude in a 6-month period? Explain, based on the probability of this occurring.

65. Load-bearing structures in underground mines are often required to carry additional loads while mining operations are in progress. As the structures adjust to this new weight, small-scale displacements take place that result in the release of seismic and acoustic energy, called *rock noise.* This energy can be detected using special geophysical equipment. Assume that in a particular mine the average number of rock noises recorded during normal activity is 3 per hour. Would you consider it unusual if more than 10 were detected in a 2-hour period? Explain, based on the probability involved.

66. A burr is a thin ridge or rough area that occurs when shaping a metal part. These must be removed by hand or by means of some newer method such as water jets, thermal energy, or electrochemical processing before the part can be used. Assume that a part used in automatic transmissions typically averages two burrs each. What is the probability that the total number of burrs found on seven randomly selected parts will be at most four?

67. Cast iron is an alloy composed primarily of iron together with smaller amounts of other elements, including carbon, silicon, sulfur, and phosphorus. The carbon occurs as graphite, which is soft, or iron carbide, which is very hard and brittle. The type of cast iron produced is determined by the amount and distribution of carbon in the iron. Five types of cast iron are identifiable. These are gray, compacted graphite, ductile, malleable, and white. In malleable cast iron the carbon is present as discrete graphite particles. Assume that in a particular casting these particles average 20 per square inch. Would it be unusual to see a 1/4-inch-square area of this casting with fewer than two graphite particles? Explain, based on the probability involved.

68. A Poisson random variable is such that it assumes the values 0 and 1 with equal probability. Find the value of the Poisson parameter k for this variable.

69. Prove Theorem 3.8.1. *Hint:* Note that

$$m_X(t) = E[e^{tX}] = \sum_{x=0}^{\infty} \frac{e^{tx}e^{-k}k^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-k}(ke^t)^x}{x!}$$

and use the Maclaurin series.

70. If the sensitivity of a motion-activated light is set correctly, the average number of times that it will be activated per week by squirrels and other small woods animals is .5. What is the average number of times that you would expect the light to be activated by these animals in a two-week period? If this occurred at least 5 times during a two-week period, would you suspect that the sensitivity needed to be adjusted? Explain based on the probability involved.

71. *Escherichia coli,* a bacterium often found in the human digestive tract, can mutate from being streptomycin sensitive to being streptomycin resistant, which can cause the individual involved to become resistant to the antibiotic streptomycin. Assume that there is an average of two streptomycin-resistant bacteria on cultures drawn from a particular patient. Each culture has an area of 80 square centimeters. What is the probability that a one-square-centimeter random sample from a single culture will contain at least one resistant bacterium? What is the probability that at least one will be found in 5 randomly selected one-square-centimeter samples? (Assume that the 5 samples are independent of one another.)

Section 3.9

72. An engine contains 5 seals that operate independently. If 3 or more seals fail, then the engine will fail. It is thought that when the temperature drops below 0° F each seal has a 10% chance of failure. Let X denote the number of seals that fail so that X is binomial with $n=5$ and $p=.10$. Simulate the performance of 10 such engines under 0° conditions. Use the 10 simulations to estimate the average number of seals that will fail per engine by averaging your 10 values of X. Compare your estimate to the theoretical mean of .5. In your simulation, how many of the 10 engines would have failed?

73. Use Table II of App. A to simulate the arrival and departure of planes to the airport described in Example 3.9.2 for 10 more 1-minute periods. Based on these data, approximate the average number of planes on the ground at a given time by finding the arithmetic average of the values of Z simulated in the experiment.

74. Consider the random variable X, the number of runs conducted to produce an unacceptable lot when coating steel tubes (see Exercise 25.) X is geometric with $p = .05$. Divide the 100 possible two-digit numbers into two categories, with numbers 00–04 denoting the production of an unacceptable lot and the remaining numbers denoting the production of an acceptable lot. Simulate the experiment of producing lots until an unacceptable one is obtained 10 times. Record the value obtained for X in each simulation. Based on these data, approximate the average value of X. Does your approximate value lie close to the theoretical mean value of 20? If not, run the simulation 10 more times. Is the arithmetic average of your observed values for X closer to 20 this time?

REVIEW EXERCISES

75. A large microprocessor chip contains multiple copies of circuits. If a circuit fails, the chip knows it and knows how to select the proper logic to repair itself. The average number of defects per chip is 300. What is the probability that 10 or fewer defects will be found in a randomly selected region that comprises 5% of the total surface area? What is the probability that more than 10 defects will be found?

76. When a program is submitted to the computer in a time-sharing system, it is processed on a space-available basis. Past experience shows that a program submitted to one such system is accepted for processing within 1 minute with probability .25. Assume that during the course of a day five programs are submitted with enough time between submissions to ensure independence. Let X denote the number of programs accepted for processing within 1 minute.
(*a*) Find $E[X]$ and Var X.
(*b*) Find the probability that none of these programs will be accepted for processing within 1 minute.
(*c*) Five programs are submitted on each of two consecutive days. What is the probability that no programs will be accepted for processing within 1 minute during this two-day period?

77. A new type of brake lining is being studied. It is thought that the lining will last for at least 70,000 miles on 90% of the cars in which it is used. Laboratory trials are conducted to simulate the driving experience of 100 cars in which this lining is used. Let X denote the number of cars whose brakes must be relined before the 70,000-mile mark.
(*a*) What is the distribution of X? What is $E[X]$?
(*b*) What distribution can be used to approximate probabilities for X?
(*c*) Suppose that we agree that the 90% figure is too high if 17 or more of the 100 cars require a relinement prior to the 70,000-mile mark. What is the probability that we will come to this conclusion by chance even though the 90% figure is correct?

78. A bank of guns fires on a target one after the other. Each has probability 1/4 of hitting the target on a given shot. Find the probability that the second hit comes before the seventh gun fires.

79. In a video game the player attempts to capture a treasure lying behind one of five doors. The location of the treasure varies randomly in such a way that at any given time it is just as likely to be behind one door as any other. When the player knocks on a given door, the treasure is his if it lies behind that door. Otherwise he must return to his original starting point and approach the doors through a dangerous maze again. Once the treasure is captured, the game ends. Let X denote the number of trials needed to capture the treasure. Find the average number of trials needed to capture the treasure. Find $P[X \leq 3]$. Find $P[X > 3]$.

80. An automobile repair shop has 10 rebuilt transmissions in stock. Three are not in correct working order and have an internal defect that will cause trouble within the first 1000 miles of operation. Four of these transmissions are randomly selected and installed in customers' cars. Find the probability that no defective transmissions are installed. Find the probability that exactly one defective transmission is installed.

81. A computer terminal can pick up an erroneous signal from the keyboard that does not show up on the screen. This creates a silent error that is difficult to detect. Assume that for a particular keyboard the probability that this will occur per entry is 1/1000. In 12,000 entries find the probability that no silent errors occur. Find the probability of at least one silent error.

82. It is thought that 1 of every 10 cars on the road has a speedometer that is miscalibrated to the extent that it reads at least 5 miles per hour low. During the course of a day 15 drivers are stopped and charged with exceeding the speed limit by at least 5 miles per hour. Would you be surprised to find that at least 5 of the cars involved have miscalibrated speedometers? Explain, based on the probability of observing a result this unusual by chance.

83. Let

$$f(x) = \frac{x^2}{14} \qquad x = 1, 2, 3$$

(*a*) Show that f is the density for a discrete random variable.
(*b*) Find $E[X]$ and $E[X^2]$ from the definition of these terms.
(*c*) Find $= m_X(t)$.
(*d*) Use $= m_X(t)$ to verify your answers to part (*b*).
(*e*) Find Var X and σ.

84. Find the expression for the cumulative distribution function for the random variable of Exercise 24. Use this function to find the probability that at least three wells must be drilled to obtain the first strike.

85. Consider the moment generating function given below. In each case, state the name of the distribution involved and the numerical value of the parameters that identify the distribution. For example, if the distribution is binomial, state the value of n and p; if geometric, give the value of p.

(*a*) $(.2 + .8e^t)^{10}$
(*b*) $e^{5(e^t - 1)}$
(*c*) $(.7 + .3e^t)$
(*d*) $\dfrac{.6e^t}{1 - .4e^t}$
(*e*) $\dfrac{(.3e^t)^5}{(1 - .7e^t)^5}$
(*f*) $e^{e^t - 1}$

86. For each of the distribution in Exercise 85, give the numerical values of the mean, variance, and standard deviation.

87. Consider the problem of Example 1.2.3. Assume that sampling is independent and that at each stage the probability of obtaining a defective part when the process is working correctly is .01. Let X denote the number of samples taken to obtain the first defective part.

(*a*) Find the density for X.

(*b*) What is the average value of X?

(*c*) What is the equation for the cumulative distribution function for X? Use F to find the probability that the first defective part will be found on or before the 90th sample.

CHAPTER

4

CONTINUOUS DISTRIBUTIONS

In Chap. 3 we learned to distinguish a discrete random variable from one that is not discrete. In this chapter we consider a large class of nondiscrete random variables. In particular, we consider random variables that are called *continuous*. We first study the general properties of variables of the continuous type and then present some important families of continuous random variables.

4.1 CONTINUOUS DENSITIES

In Chap. 3 we considered the random variable T, the time of the peak demand for electricity at a particular power plant. We agreed that this random variable is not discrete since, "a priori"—before the fact—we cannot limit the set of possible values for T to some finite or countably infinite collection of times. Time is measured continuously, and T can conceivably assume any value in the time interval $[0, 24)$, where 0 denotes 12 midnight one day and 24 denotes 12 midnight the next day. Furthermore, if we ask *before* the day begins, What is the probability that the peak demand will occur exactly 12.013 278 650 931 271? the answer is 0. It is virtually impossible for the peak load to occur at this split second in time, not the slightest bit earlier or later. These two properties, possible values occurring as intervals and the a priori probability of assuming any specific value being 0, are the characteristics that identify a random variable as being continuous. This leads us to our next definition.

Definition 4.1.1 (Continuous random variable). A random variable is continuous if it can assume any value in some interval or intervals of real numbers and the probability that it assumes any specific value is 0.