Chapter

2

DISCRETE DISTRIBUTIONS

2.1 RANDOM VARIABLES OF THE DISCRETE TYPE

An outcome space S may be difficult to describe if the elements of S are not numbers. We shall now discuss how we can use a rule by which each outcome of a random experiment, an element s of S, may be associated with a real number x. We begin the discussion with an example.

Example 2.1-1 A rat is selected at random from a cage and its sex is determined. The set of possible outcomes is female and male. Thus, the outcome space is $S = \{\text{female, male}\} = \{\text{F, M}\}$. Let X be a function defined on S such that $X(\text{F}) = 0$ and $X(\text{M}) = 1$. X is then a real-valued function that has the outcome space S as its domain and the set of real numbers $\{x : x = 0, 1\}$ as its range. We call X a random variable, and in this example, the space associated with X is the set of numbers $\{x : x = 0, 1\}$. ■

We now formulate the definition of a random variable.

Definition 2.1-1
Given a random experiment with an outcome space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a **random variable**. The **space** of X is the set of real numbers $\{x : X(s) = x, s \in S\}$, where $s \in S$ means that the element s belongs to the set S.

REMARK As we give examples of random variables and their probability distributions, the reader will soon recognize that, when observing a random experiment, the experimenter must take some type of measurement (or measurements). This measurement can be thought of as the outcome of a random variable. We would simply like to know the probability of a measurement resulting in A, a subset of the space of X. If this is known for all subsets A, then we know the probability distribution of the random variable. Obviously, in practice, we often do not know this distribution exactly. Hence, statisticians make conjectures about these distributions; that is, we construct probabilistic models for random variables. The ability of a statistician to model a real situation appropriately is a valuable trait. In this chapter we introduce some probability models in which the spaces of the random variables consist of sets of integers. ■

It may be that the set S has elements that are themselves real numbers. In such an instance, we could write $X(s) = s$, so that X is the identity function and the space of X is also S. This situation is illustrated in Example 2.1-2.

Example 2.1-2 Let the random experiment be the cast of a die. Then the outcome space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with the elements of S indicating the number of spots on the side facing up. For each $s \in S$, let $X(s) = s$. The space of the random variable X is then $\{1, 2, 3, 4, 5, 6\}$.

If we associate a probability of 1/6 with each outcome, then, for example, $P(X = 5) = 1/6$, $P(2 \le X \le 5) = 4/6$, and $P(X \le 2) = 2/6$ seem to be reasonable assignments, where, in this example, $\{2 \le X \le 5\}$ means $\{X = 2, 3, 4, \text{ or } 5\}$ and $\{X \le 2\}$ means $\{X = 1 \text{ or } 2\}$. ■

The student will no doubt recognize two major difficulties here:

1. In many practical situations, the probabilities assigned to the events are unknown.
2. Since there are many ways of defining a function X on S, which function do we want to use?

As a matter of fact, the solutions to these problems in particular cases are major concerns in applied statistics. In considering (2), statisticians try to determine what *measurement* (or measurements) should be taken on an outcome; that is, how best do we "mathematize" the outcome? These measurement problems are most difficult and can be answered only by getting involved in a practical project. For (1), we often need to estimate these probabilities or percentages through repeated observations (called sampling). For example, what percentage of newborn girls in the University of Iowa Hospital weigh less than 7 pounds? Here a newborn baby girl is the outcome, and we have measured her one way (by weight), but obviously there are many other ways of measuring her. If we let X be the weight in pounds, we are interested in the probability $P(X < 7)$, and we can estimate this probability only by repeated observations. One obvious way of estimating it is by the use of the relative frequency of $\{X < 7\}$ after a number of observations. If it is reasonable to make additional assumptions, we will study other ways of estimating that probability. It is this latter aspect with which the field of mathematical statistics is concerned. That is, if we assume certain models, we find that the theory of statistics can explain how best to draw conclusions or make predictions.

In many instances, it is clear exactly what function X the experimenter wants to define on the outcome space. For example, the caster in the dice game called craps is concerned about the sum of the spots (say X) that are facing upward on the pair of dice. Hence, we go directly to the space of X, which we shall denote by the same letter S. After all, in the dice game the caster is directly concerned only with the probabilities associated with X. Thus, for convenience, in many instances the reader can think of the space of X as being the outcome space.

Let X denote a random variable with space S. Suppose that we know how the probability is distributed over the various subsets A of S; that is, we can compute $P(X \in A)$. In this sense, we speak of the distribution of the random variable X, meaning, of course, the distribution of probability associated with the space S of X.

Let X denote a random variable with one-dimensional space S, a subset of the real numbers. Suppose that the space S contains a countable number of points; that is, either S contains a finite number of points, or the points of S can be put into a one-to-one correspondence with the positive integers. Such a set S is called a set of discrete points or simply a discrete outcome space. Furthermore, any random variable defined on such an S can assume at most a countable number of values, and is therefore called a random variable of the **discrete type**. The corresponding probability distribution likewise is said to be of the discrete type.

For a random variable X of the discrete type, the probability $P(X = x)$ is frequently denoted by $f(x)$, and this function $f(x)$ is called the **probability mass function**. Note that some authors refer to $f(x)$ as the probability function, the frequency function, or the probability density function. In the discrete case, we shall use "probability mass function," and it is hereafter abbreviated pmf.

Let $f(x)$ be the pmf of the random variable X of the discrete type, and let S be the space of X. Since $f(x) = P(X = x)$ for $x \in S$, $f(x)$ must be nonnegative for $x \in S$, and we want all these probabilities to add to 1 because each $P(X = x)$ represents the fraction of times x can be expected to occur. Moreover, to determine the probability associated with the event $A \in S$, we would sum the probabilities of the x values in A. This leads us to the following definition.

Definition 2.1-2

The pmf $f(x)$ of a discrete random variable X is a function that satisfies the following properties:

(a) $f(x) > 0, \qquad x \in S;$

(b) $\displaystyle\sum_{x \in S} f(x) = 1;$

(c) $\displaystyle P(X \in A) = \sum_{x \in A} f(x), \qquad \text{where } A \subset S.$

Of course, we usually let $f(x) = 0$ when $x \notin S$; thus, the domain of $f(x)$ is the set of real numbers. When we define the pmf $f(x)$ and do not say "zero elsewhere," we tacitly mean that $f(x)$ has been defined at all x's in the space S and it is assumed that $f(x) = 0$ elsewhere; that is, $f(x) = 0$ when $x \notin S$. Since the probability $P(X = x) = f(x) > 0$ when $x \in S$, and since S contains all the outcomes with positive probabilities associated with X, we sometimes refer to S as the **support** of X as well as the space of X.

Cumulative probabilities are often of interest. We call the function defined by

$$F(x) = P(X \le x), \qquad -\infty < x < \infty,$$

the **cumulative distribution function** and abbreviate it as cdf. The cdf is sometimes referred to as the **distribution function** of the random variable X. Values of the cdf of certain random variables are given in the appendix and will be pointed out as we use them (see Appendix B, Tables II, III, IV, Va, VI, VII, and IX).

When a pmf is constant on the space or support, we say that the distribution is **uniform** over that space. As an illustration, in Example 2.1-2 X has a discrete uniform distribution on $S = \{1, 2, 3, 4, 5, 6\}$ and its pmf is

$$f(x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6.$$

We can generalize this result by letting X have a discrete uniform distribution over the first m positive integers, so that its pmf is

$$f(x) = \frac{1}{m}, \qquad x = 1, 2, 3, \ldots, m.$$

The cdf of X is defined as follows where $k = 1, 2, \ldots, m-1$. We have

$$F(x) = P(X \le x) = \begin{cases} 0, & x < 1, \\ \dfrac{k}{m}, & k \le x < k+1, \\ 1, & m \le x. \end{cases}$$

Note that this is a step function with a jump of size $1/m$ for $x = 1, 2, \ldots, m$.

We now give an example in which X does not have a uniform distribution.

Example 2.1-3 Roll a fair four-sided die twice, and let X be the maximum of the two outcomes. The outcome space for this experiment is $S_0 = \{(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}$, where we assume that each of these 16 points has probability 1/16. Then $P(X = 1) = P[(1, 1)] = 1/16$, $P(X = 2) = P[\{(1, 2), (2, 1), (2, 2)\}] = 3/16$, and similarly $P(X = 3) = 5/16$ and $P(X = 4) = 7/16$. That is, the pmf of X can be written simply as

$$f(x) = P(X = x) = \frac{2x-1}{16}, \qquad x = 1, 2, 3, 4. \tag{2.1-1}$$

We could add that $f(x) = 0$ elsewhere; but if we do not, the reader should take $f(x)$ to equal zero when $x \notin S = \{1, 2, 3, 4\}$. ■

A better understanding of a particular probability distribution can often be obtained with a graph that depicts the pmf of X. Note that the graph of the pmf when $f(x) > 0$ would be simply the set of points $\{[x, f(x)] : x \in S\}$, where S is the space of X. Two types of graphs can be used to give a better visual appreciation of the pmf: a line graph and a probability histogram. A **line graph** of the pmf $f(x)$ of the random variable X is a graph having a vertical line segment drawn from $(x, 0)$ to $[x, f(x)]$ at each x in S, the space of X. If X can assume only integer values, a **probability histogram** of the pmf $f(x)$ is a graphical representation that has a rectangle of height $f(x)$ and a base of length 1, centered at x for each $x \in S$, the space of X. Thus, the area of each rectangle is equal to the respective probability $f(x)$, and the total area of a probability histogram is 1.

Figure 2.1-1 displays a line graph and a probability histogram for the pmf $f(x)$ defined in Equation 2.1-1.

Our next probability model uses the material in Section 1.2 on methods of enumeration. Consider a collection of $N = N_1+N_2$ similar objects, N_1 of them belonging to one of two dichotomous classes (red chips, say) and N_2 of them belonging to the second class (blue chips, say). A collection of n objects is selected from these N objects at random and without replacement. Find the probability that exactly x (where the nonnegative integer x satisfies $x \le n$, $x \le N_1$, and $n - x \le N_2$) of these n objects belong to the first class and $n - x$ belong to the second. Of course, we can select x objects from the first class in any one of $\binom{N_1}{x}$ ways and $n-x$ objects from the second class in any one of $\binom{N_2}{n-x}$ ways. By the multiplication principle, the product $\binom{N_1}{x}\binom{N_2}{n-x}$ equals the number of ways the joint operation can be performed.

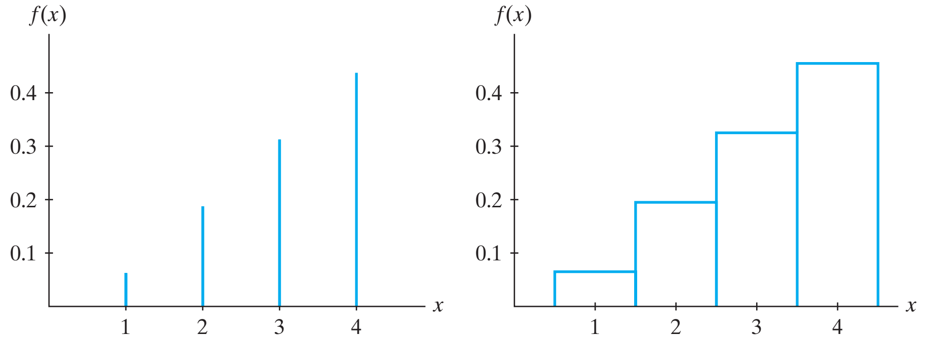

Figure 2.1-1 Line graph and probability histogram

If we assume that each of the $\binom{N}{n}$ ways of selecting n objects from $N = N_1 + N_2$ objects has the same probability, it follows that the desired probability is

$$f(x) = P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}},$$

where the space S is the collection of nonnegative integers x that satisfies the inequalities $x \le n$, $x \le N_1$, and $n - x \le N_2$. We say that the random variable X has a **hypergeometric distribution**.

Example 2.1-4 Some examples of hypergeometric probability histograms are given in Figure 2.1-2. The values of N_1, N_2, and n are given with each figure. ■

Example 2.1-5 In a small pond there are 50 fish, 10 of which have been tagged. If a fisherman's catch consists of 7 fish selected at random and without replacement, and X denotes the number of tagged fish, the probability that exactly 2 tagged fish are caught is

$$P(X = 2) = \frac{\binom{10}{2}\binom{40}{5}}{\binom{50}{7}} = \frac{(45)(658{,}008)}{99{,}884{,}400} = \frac{246{,}753}{832{,}370} = 0.2964.$$ ■

Example 2.1-6 A lot (collection) consisting of 100 fuses is inspected by the following procedure: Five fuses are chosen at random and tested; if all five blow at the correct amperage, the lot is accepted. Suppose that the lot contains 20 defective fuses. If X is a random variable equal to the number of defective fuses in the sample of 5, the probability of accepting the lot is

$$P(X = 0) = \frac{\binom{20}{0}\binom{80}{5}}{\binom{100}{5}} = \frac{19{,}513}{61{,}110} = 0.3193.$$

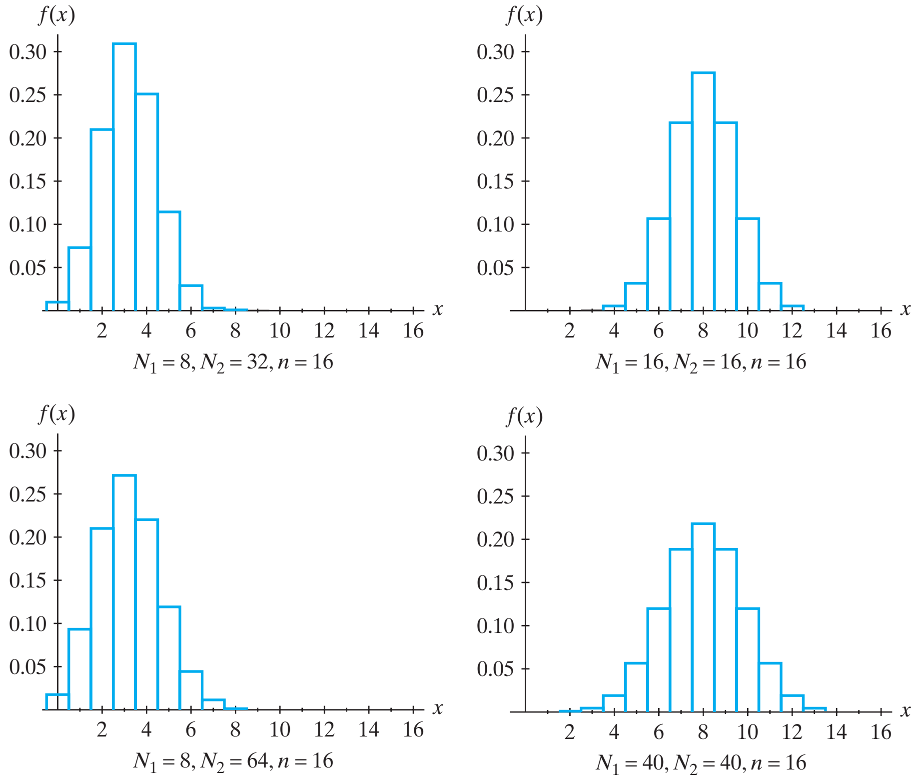

Figure 2.1-2 Hypergeometric probability histograms

More generally, the pmf of X is

$$f(x) = P(X = x) = \frac{\dbinom{20}{x}\dbinom{80}{5-x}}{\dbinom{100}{5}}, \qquad x = 0, 1, 2, 3, 4, 5.$$

In Section 1.1, we discussed the relationship between the probability $P(A)$ of an event A and the relative frequency $\mathcal{N}(A)/n$ of occurrences of event A in n repetitions of an experiment. We shall now extend those ideas.

Suppose that a random experiment is repeated n independent times. Let $A = \{X = x\}$, the event that x is the outcome of the experiment. Then we would expect the relative frequency $\mathcal{N}(A)/n$ to be close to $f(x)$. The next example illustrates this property.

Example 2.1-7 A fair four-sided die with outcomes 1, 2, 3, and 4 is rolled twice. Let X equal the sum of the two outcomes. Then the possible values of X are 2, 3, 4, 5, 6, 7, and 8. The following argument suggests that the pmf of X is given by $f(x) = (4 - |x - 5|)/16$, for $x = 2, 3, 4, 5, 6, 7, 8$ [i.e., $f(2) = 1/16$, $f(3) = 2/16$, $f(4) = 3/16$, $f(5) = 4/16$, $f(6) = 3/16$, $f(7) = 2/16$, and $f(8) = 1/16$]: Intuitively, these probabilities seem correct if we think of the 16 points (result on first roll, result on second roll) and

Table 2.1-1 Sum of two tetrahedral dice

x	Number of Observations of x	Relative Frequency of x	Probability of $\{X = x\}, f(x)$
2	71	0.071	0.0625
3	124	0.124	0.1250
4	194	0.194	0.1875
5	258	0.258	0.2500
6	177	0.177	0.1875
7	122	0.122	0.1250
8	54	0.054	0.0625

assume that each has probability 1/16. Then note that $X = 2$ only for the point (1, 1), $X = 3$ for the two points (2, 1) and (1, 2), and so on. This experiment was simulated 1000 times on a computer. Table 2.1-1 lists the results and compares the relative frequencies with the corresponding probabilities.

A graph can be used to display the results shown in Table 2.1-1. The probability histogram of the pmf $f(x)$ of X is given by the dotted lines in Figure 2.1-3. It is superimposed over the shaded histogram that represents the observed relative frequencies of the corresponding x values. The shaded histogram is the **relative frequency histogram**. For random experiments of the discrete type, this relative frequency histogram of a set of data gives an estimate of the probability histogram of the associated random variable when the latter is unknown. (Estimation is considered in detail later in the book.) ■

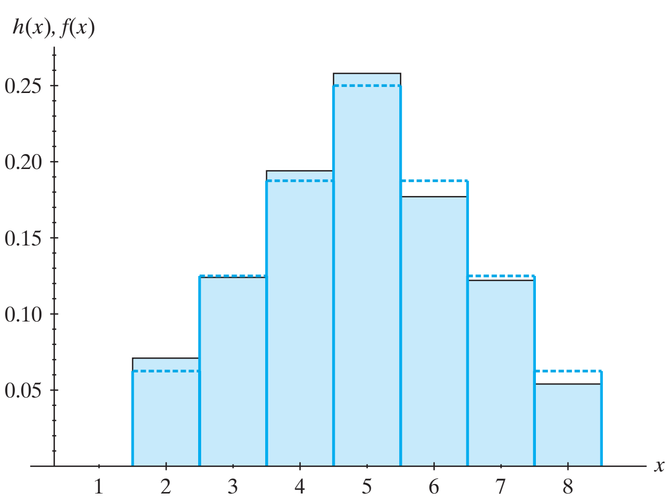

Figure 2.1-3 Sum of two tetrahedral dice

Exercises

2.1-1. Let the pmf of X be defined by $f(x) = x/9$, $x = 2, 3, 4$.

(a) Draw a line graph for this pmf.

(b) Draw a probability histogram for this pmf.

2.1-2. Let a chip be taken at random from a bowl that contains six white chips, three red chips, and one blue chip. Let the random variable $X = 1$ if the outcome is a white chip, let $X = 5$ if the outcome is a red chip, and let $X = 10$ if the outcome is a blue chip.

(a) Find the pmf of X.

(b) Graph the pmf as a line graph.

2.1-3. For each of the following, determine the constant c so that $f(x)$ satisfies the conditions of being a pmf for a random variable X, and then depict each pmf as a line graph:

(a) $f(x) = x/c, \qquad x = 1, 2, 3, 4.$

(b) $f(x) = cx, \qquad x = 1, 2, 3, \ldots, 10.$

(c) $f(x) = c(1/4)^x, \qquad x = 1, 2, 3, \ldots.$

(d) $f(x) = c(x+1)^2, \qquad x = 0, 1, 2, 3.$

(e) $f(x) = x/c, \qquad x = 1, 2, 3, \ldots, n.$

(f) $f(x) = \dfrac{c}{(x+1)(x+2)}, \qquad x = 0, 1, 2, 3, \ldots.$

HINT: In part (f), write $f(x) = 1/(x+1) - 1/(x+2)$.

2.1-4. The state of Michigan generates a three-digit number at random twice a day, seven days a week for its Daily 3 game. The numbers are generated one digit at a time. Consider the following set of 50 three-digit numbers as 150 one-digit integers that were generated at random:

169	938	506	757	594	656	444	809	321	545
732	146	713	448	861	612	881	782	209	752
571	701	852	924	766	633	696	023	601	789
137	098	534	826	642	750	827	689	979	000
933	451	945	464	876	866	236	617	418	988

Let X denote the outcome when a single digit is generated.

(a) With true random numbers, what is the pmf of X? Draw the probability histogram.

(b) For the 150 observations, determine the relative frequencies of 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively.

(c) Draw the relative frequency histogram of the observations on the same graph paper as that of the probability histogram. Use a colored or dashed line for the relative frequency histogram.

2.1-5. The pmf of X is $f(x) = (5-x)/10, \; x = 1, 2, 3, 4$.

(a) Graph the pmf as a line graph.

(b) Use the following independent observations of X, simulated on a computer, to construct a table like Table 2.1-1:

3	1	2	2	3	2	2	2	1	3	3	2	3	2	4	4	2	1	1	3
3	1	2	2	1	1	4	2	3	1	1	1	2	1	3	1	1	3	3	1
1	1	1	1	1	4	1	3	1	2	4	1	1	2	3	4	3	1	4	2
2	1	3	2	1	4	1	1	1	2	1	3	4	3	2	1	4	4	1	3
2	2	2	1	2	3	1	1	4	2	1	4	2	1	2	3	1	4	2	3

(c) Construct a probability histogram and a relative frequency histogram like Figure 2.1-3.

2.1-6. Let a random experiment be the casting of a pair of fair dice, each having six faces, and let the random variable X denote the sum of the dice.

(a) With reasonable assumptions, determine the pmf $f(x)$ of X. HINT: Picture the sample space consisting of the 36 points (result on first die, result on second die), and assume that each has probability 1/36. Find the probability of each possible outcome of X, namely, $x = 2, 3, 4, \ldots, 12$.

(b) Draw a probability histogram for $f(x)$.

2.1-7. Let a random experiment be the casting of a pair of fair six-sided dice and let X equal the minimum of the two outcomes.

(a) With reasonable assumptions, find the pmf of X.

(b) Draw a probability histogram of the pmf of X.

(c) Let Y equal the range of the two outcomes (i.e., the absolute value of the difference of the largest and the smallest outcomes). Determine the pmf $g(y)$ of Y for $y = 0, 1, 2, 3, 4, 5$.

(d) Draw a probability histogram for $g(y)$.

2.1-8. A fair four-sided die has two faces numbered 0 and two faces numbered 2. Another fair four-sided die has its faces numbered 0, 1, 4, and 5. The two dice are rolled. Let X and Y be the respective outcomes of the roll. Let $W = X + Y$.

(a) Determine the pmf of W.

(b) Draw a probability histogram of the pmf of W.

2.1-9. The pmf of X is $f(x) = (1 + |x - 3|)/11$, for $x = 1, 2, 3, 4, 5$. Graph this pmf as a line graph.

2.1-10. Suppose there are 3 defective items in a lot (collection) of 50 items. A sample of size 10 is taken at random and without replacement. Let X denote the number of defective items in the sample. Find the probability that the sample contains

(a) Exactly one defective item.

(b) At most one defective item.

2.1-11. In a lot (collection) of 100 light bulbs, there are 5 defective bulbs. An inspector inspects 10 bulbs selected at random. Find the probability of finding at least one defective bulb. HINT: First compute the probability of finding no defectives in the sample.

2.1-12. Let X be the number of accidents per week in a factory. Let the pmf of X be

$$f(x) = \frac{1}{(x+1)(x+2)} = \frac{1}{x+1} - \frac{1}{x+2}, \qquad x = 0, 1, 2, \ldots.$$

Find the conditional probability of $X \geq 4$, given that $X \geq 1$.

2.1-13. A professor gave her students six essay questions from which she will select three for a test. A student has time to study for only three of these questions. What is the probability that, of the questions studied,

(a) at least one is selected for the test?

(b) all three are selected?

(c) exactly two are selected?

2.1-14. Often in buying a product at a supermarket, there is a concern about the item being underweight. Suppose there are 20 "one-pound" packages of frozen ground turkey on display and 3 of them are underweight. A consumer group buys 5 of the 20 packages at random. What is the probability of at least one of the five being underweight?

2.1-15. Five cards are selected at random without replacement from a standard, thoroughly shuffled 52-card deck of playing cards. Let X equal the number of face cards (kings, queens, jacks) in the hand. Forty observations of X yielded the following data:

2 1 2 1 0 0 1 0 1 1 0 2 0 2 3 0 1 1 0 3

1 2 0 2 0 2 0 1 0 1 1 2 1 0 1 1 2 1 1 0

(a) Argue that the pmf of X is

$$f(x) = \frac{\binom{12}{x}\binom{40}{5-x}}{\binom{52}{5}}, \qquad x = 0, 1, 2, 3, 4, 5,$$

and thus, that $f(0) = 2109/8330$, $f(1) = 703/1666$, $f(2) = 209/833$, $f(3) = 55/833$, $f(4) = 165/21{,}658$, and $f(5) = 33/108{,}290$.

(b) Draw a probability histogram for this distribution.

(c) Determine the relative frequencies of $0, 1, 2, 3$, and superimpose the relative frequency histogram on your probability histogram.

2.1-16. (Michigan Mathematics Prize Competition, 1992, Part II) From the set $\{1, 2, 3, \ldots, n\}$, k distinct integers are selected at random and arranged in numerical order (from lowest to highest). Let $P(i, r, k, n)$ denote the probability that integer i is in position r. For example, observe that $P(1, 2, k, n) = 0$, as it is impossible for the number 1 to be in the second position after ordering.

(a) Compute $P(2, 1, 6, 10)$.

(b) Find a general formula for $P(i, r, k, n)$.

2.1-17. A bag contains 144 ping-pong balls. More than half of the balls are painted orange and the rest are painted blue. Two balls are drawn at random without replacement. The probability of drawing two balls of the same color is the same as the probability of drawing two balls of different colors. How many orange balls are in the bag?

2.2 MATHEMATICAL EXPECTATION

An extremely important concept in summarizing important characteristics of distributions of probability is that of mathematical expectation, which we introduce with an example.

Example 2.2-1

An enterprising young man who needs a little extra money devises a game of chance in which some of his friends might wish to participate. The game that he proposes is to let the participant cast a fair die and then receive a payment according to the following schedule: If the event $A = \{1, 2, 3\}$ occurs, he receives one dollar; if $B = \{4, 5\}$ occurs, he receives two dollars; and if $C = \{6\}$ occurs, he receives three dollars. If X is a random variable that represents the payoff, then the pmf of X is given by

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

that is, $f(1) = 3/6$, $f(2) = 2/6$, $f(3) = 1/6$. If the game is repeated a large number of times, the payment of one dollar would occur about 3/6 of the times, two dollars about 2/6 of the times, and three dollars about 1/6 of the times. Thus, the average payment would be

$$(1)\left(\frac{3}{6}\right) + (2)\left(\frac{2}{6}\right) + (3)\left(\frac{1}{6}\right) = \frac{10}{6} = \frac{5}{3}.$$

That is, the young man expects to pay 5/3 of a dollar "on the average." This is called the mathematical expectation of the payment. If the young man could charge two dollars to play the game, he could make $2 - 5/3 = 1/3$ of a dollar on the average each play. Note that this mathematical expectation can be written

$$E(X) = \sum_{x=1}^{3} x f(x)$$

and is often denoted by the Greek letter μ, which is called the mean of X or of its distribution. ■

Suppose that we are interested in another function of X, say $u(X)$. Let us call it $Y = u(X)$. Of course, Y is a random variable and has a pmf. For illustration, in Example 2.2-1, $Y = X^2$ has the pmf

$$g(y) = (4 - \sqrt{y}\,)/6, \qquad y = 1, 4, 9;$$

that is, $g(1) = 3/6$, $g(4) = 2/6$, $g(9) = 1/6$. Moreover, where S_Y is the space of Y, the mean of Y is

$$\mu_Y = \sum_{y \in S_Y} y\, g(y) = (1)\left(\frac{3}{6}\right) + (4)\left(\frac{2}{6}\right) + (9)\left(\frac{1}{6}\right) = \frac{20}{6} = \frac{10}{3}.$$

Participants in the young man's game might be more willing to play this game for 4 dollars as they can win $9 - 4 = 5$ dollars and lose only $4 - 1 = 3$ dollars. Note that the young man can expect to win $4 - 10/3 = 2/3$ of a dollar on the average each play. A game based upon $Z = X^3$ might even be more attractive to participants if the young man charges 10 dollars to play this game. Then the participant could win $27 - 10 = 17$ dollars and lose only $10 - 1 = 9$ dollars. The details of this latter game are covered in Exercise 2.2-5.

In any case, it is important to note that

$$E(Y) = \sum_{y \in S_Y} y\, g(y) = \sum_{x \in S_X} x^2 f(x) = \frac{20}{6} = \frac{10}{3}.$$

That is, the same value is obtained by either formula. While we have not proved, for a general function $u(x)$, that if $Y = u(X)$, then

$$\sum_{y \in S_Y} y\, g(y) = \sum_{x \in S_X} u(x) f(x);$$

we have illustrated it in this simple case. This discussion suggests the more general definition of mathematical expectation of a function of X.

Definition 2.2-1
If $f(x)$ is the pmf of the random variable X of the discrete type with space S, and if the summation

$$\sum_{x\in S} u(x)f(x), \qquad \text{which is sometimes written} \qquad \sum_{S} u(x)f(x),$$

exists, then the sum is called the **mathematical expectation** or the **expected value** of $u(X)$, and it is denoted by $E[u(X)]$. That is,

$$E[u(X)] = \sum_{x\in S} u(x)f(x).$$

We can think of the expected value $E[u(X)]$ as a weighted mean of $u(x)$, $x \in S$, where the weights are the probabilities $f(x) = P(X = x)$, $x \in S$.

REMARK The usual definition of mathematical expectation of $u(X)$ requires that the sum converge absolutely—that is, that

$$\sum_{x\in S} |u(x)|\, f(x)$$

converge and be finite. The reason for the absolute convergence is that it allows one, in the advanced proof of

$$\sum_{x\in S_X} u(x)f(x) = \sum_{y\in S_Y} yg(y),$$

to rearrange the order of the terms in the x-summation. In this book, each $u(x)$ is such that the convergence is absolute. ■

We provide another example.

Example 2.2-2 Let the random variable X have the pmf

$$f(x) = \frac{1}{3}, \qquad x \in S_X,$$

where $S_X = \{-1, 0, 1\}$. Let $u(X) = X^2$. Then

$$E(X^2) = \sum_{x\in S_X} x^2 f(x) = (-1)^2\left(\frac{1}{3}\right) + (0)^2\left(\frac{1}{3}\right) + (1)^2\left(\frac{1}{3}\right) = \frac{2}{3}.$$

However, the support of the random variable $Y = X^2$ is $S_Y = \{0, 1\}$ and

$$\begin{aligned} P(Y = 0) = P(X = 0) &= \frac{1}{3}, \\ P(Y = 1) = P(X = -1) + P(X = 1) &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

That is,

$$g(y) = \begin{cases} \dfrac{1}{3}, & y = 0, \\[2ex] \dfrac{2}{3}, & y = 1; \end{cases}$$

and $S_Y = \{0, 1\}$. Hence,

$$\mu_Y = E(Y) = \sum_{y \in S_Y} y\, g(y) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{2}{3}\right) = \frac{2}{3},$$

which again illustrates the preceding observation. ■

Before presenting additional examples, we list some useful facts about mathematical expectation in the following theorem.

Theorem 2.2-1 When it exists, the mathematical expectation E satisfies the following properties:

(a) If c is a constant, then $E(c) = c$.

(b) If c is a constant and u is a function, then

$$E[c\, u(X)] = cE[u(X)].$$

(c) If c_1 and c_2 are constants and u_1 and u_2 are functions, then

$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)].$$

Proof First, for the proof of (a), we have

$$E(c) = \sum_{x \in S} c f(x) = c \sum_{x \in S} f(x) = c$$

because

$$\sum_{x \in S} f(x) = 1.$$

Next, to prove (b), we see that

$$\begin{aligned} E[c\, u(X)] &= \sum_{x \in S} c\, u(x) f(x) \\ &= c \sum_{x \in S} u(x) f(x) \\ &= c\, E[u(X)]. \end{aligned}$$

Finally, the proof of (c) is given by

$$\begin{aligned} E[c_1 u_1(X) + c_2 u_2(X)] &= \sum_{x \in S} [c_1 u_1(x) + c_2 u_2(x)] f(x) \\ &= \sum_{x \in S} c_1 u_1(x) f(x) + \sum_{x \in S} c_2 u_2(x) f(x). \end{aligned}$$

By applying (b), we obtain

$$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)].$$ □

Property (c) can be extended to more than two terms by mathematical induction; that is, we have

$$\textbf{(c}'\textbf{)}\quad E\left[\sum_{i=1}^{k} c_i\, u_i(X)\right] = \sum_{i=1}^{k} c_i\, E[u_i(X)].$$

Because of property (c'), the mathematical expectation E is often called a **linear** or **distributive** operator.

Example 2.2-3 Let X have the pmf

$$f(x) = \frac{x}{10}, \qquad x = 1, 2, 3, 4.$$

Then the mean of X is

$$\begin{aligned}
\mu = E(X) &= \sum_{x=1}^{4} x\left(\frac{x}{10}\right) \\
&= (1)\left(\frac{1}{10}\right) + (2)\left(\frac{2}{10}\right) + (3)\left(\frac{3}{10}\right) + (4)\left(\frac{4}{10}\right) = 3,
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \sum_{x=1}^{4} x^2\left(\frac{x}{10}\right) \\
&= (1)^2\left(\frac{1}{10}\right) + (2)^2\left(\frac{2}{10}\right) + (3)^2\left(\frac{3}{10}\right) + (4)^2\left(\frac{4}{10}\right) = 10,
\end{aligned}$$

and

$$E[X(5-X)] = 5E(X) - E(X^2) = (5)(3) - 10 = 5.$$

■

Example 2.2-4 Let $u(x) = (x-b)^2$, where b is not a function of X, and suppose $E[(X-b)^2]$ exists. To find that value of b for which $E[(X-b)^2]$ is a minimum, we write

$$\begin{aligned}
g(b) = E[(X-b)^2] &= E[X^2 - 2bX + b^2] \\
&= E(X^2) - 2bE(X) + b^2
\end{aligned}$$

because $E(b^2) = b^2$. To find the minimum, we differentiate $g(b)$ with respect to b, set $g'(b) = 0$, and solve for b as follows:

$$\begin{aligned}
g'(b) &= -2E(X) + 2b = 0, \\
b &= E(X).
\end{aligned}$$

Since $g''(b) = 2 > 0$, the mean of X, $\mu = E(X)$, is the value of b that minimizes $E[(X-b)^2]$.

■

Example 2.2-5 Let X have a hypergeometric distribution in which n objects are selected from $N = N_1 + N_2$ objects as described in Section 2.1. Then

$$\mu = E(X) = \sum_{x \in S} x\, \frac{\dbinom{N_1}{x}\dbinom{N_2}{n-x}}{\dbinom{N}{n}}.$$

Since the first term of this summation equals zero when $x = 0$, and since

$$\binom{N}{n} = \left(\frac{N}{n}\right)\binom{N-1}{n-1},$$

we can write

$$E(X) = \sum_{0<x\in S} x \frac{N_1!}{x!(N_1-x)!} \frac{\binom{N_2}{n-x}}{\left(\frac{N}{n}\right)\binom{N-1}{n-1}}.$$

Of course, $x/x! = 1/(x-1)!$ when $x \neq 0$; thus,

$$E(X) = \left(\frac{n}{N}\right) \sum_{0<x\in S} \frac{(N_1)(N_1-1)!}{(x-1)!(N_1-x)!} \frac{\binom{N_2}{n-x}}{\binom{N-1}{n-1}}$$

$$= n\left(\frac{N_1}{N}\right) \sum_{0<x\in S} \frac{\binom{N_1-1}{x-1}\binom{N_2}{n-1-(x-1)}}{\binom{N-1}{n-1}}.$$

However, when $x > 0$, the summand of this last expression represents the probability of obtaining, say, $x-1$ red chips if $n-1$ chips are selected from $N_1 - 1$ red chips and N_2 blue chips. Since the summation is over all possible values of $x-1$, it must sum to 1, as it is the sum of all possible probabilities of $x-1$. Thus,

$$\mu = E(X) = n\left(\frac{N_1}{N}\right),$$

which is a result that agrees with our intuition: We expect the number X of red chips to equal the product of the number n of selections and the fraction N_1/N of red chips in the original collection. ■

Example 2.2-6

Say an experiment has probability of success p, where $0 < p < 1$, and probability of failure $1 - p = q$. This experiment is repeated independently until the first success occurs; say this happens on the X trial. Clearly the space of X is $S_X = \{1, 2, 3, 4, \ldots\}$. What is $P(X = x)$, where $x \in S_X$? We must observe $x-1$ failures and then a success to have this happen. Thus, due to the independence, the probability is

$$f(x) = P(X = x) = \overbrace{q \cdot q \cdots q}^{x-1\, q's} \cdot p = q^{x-1}p, \qquad x \in S_X.$$

Since p and q are positive, this is a pmf because

$$\sum_{x\in S_X} q^{x-1}p = p(1 + q + q^2 + q^3 + \cdots) = \frac{p}{1-q} = \frac{p}{p} = 1.$$

The mean of this **geometric distribution** is

$$\mu = \sum_{x=1}^{\infty} xf(x) = (1)p + (2)qp + (3)q^2p + \cdots$$

and

$$q\mu = (q)p + (2)q^2p + (3)q^3p + \cdots .$$

If we subtract the second of these two equations from the first, we have

$$\begin{aligned}(1-q)\mu &= p + pq + pq^2 + pq^3 + \cdots \\ &= (p)(1 + q + q^2 + q^3 + \cdots) \\ &= (p)\left(\frac{1}{1-q}\right) = 1.\end{aligned}$$

That is,

$$\mu = \frac{1}{1-q} = \frac{1}{p}.$$

For illustration, if $p = 1/10$, we would expect $\mu = 10$ trials are needed on the average to observe a success. This certainly agrees with our intuition. ■

Exercises

2.2-1. Find $E(X)$ for each of the distributions given in Exercise 2.1-3.

2.2-2. Let the random variable X have the pmf

$$f(x) = \frac{(|x|+1)^2}{9}, \qquad x = -1, 0, 1.$$

Compute $E(X)$, $E(X^2)$, and $E(3X^2 - 2X + 4)$.

2.2-3. Let the random variable X be the number of days that a certain patient needs to be in the hospital. Suppose X has the pmf

$$f(x) = \frac{5-x}{10}, \qquad x = 1, 2, 3, 4.$$

If the patient is to receive \$200 from an insurance company for each of the first two days in the hospital and \$100 for each day after the first two days, what is the expected payment for the hospitalization?

2.2-4. An insurance company sells an automobile policy with a deductible of one unit. Let X be the amount of the loss having pmf

$$f(x) = \begin{cases} 0.9, & x = 0, \\ \dfrac{c}{x}, & x = 1, 2, 3, 4, 5, 6, \end{cases}$$

where c is a constant. Determine c and the expected value of the amount the insurance company must pay.

2.2-5. In Example 2.2-1 let $Z = u(X) = X^3$.

(a) Find the pmf of Z, say $h(z)$.

(b) Find $E(Z)$.

(c) How much, on average, can the young man expect to win on each play if he charges \$10 per play?

2.2-6. Let the pmf of X be defined by $f(x) = 6/(\pi^2 x^2)$, $x = 1, 2, 3, \ldots$. Show that $E(X) = +\infty$ and thus, does not exist.

2.2-7. In the gambling game chuck-a-luck, for a \$1 bet it is possible to win \$1, \$2, or \$3 with respective probabilities 75/216, 15/216, and 1/216. One dollar is lost with probability 125/216. Let X equal the payoff for this game and find $E(X)$. Note that when a bet is won, the \$1 that was bet, in addition to the \$1, \$2, or \$3 that is won, is returned to the bettor.

2.2-8. Let X be a random variable with support $\{1, 2, 3, 5, 15, 25, 50\}$, each point of which has the same probability 1/7. Argue that $c = 5$ is the value that minimizes $h(c) = E(|X - c|)$. Compare c with the value of b that minimizes $g(b) = E[(X - b)^2]$.

2.2-9. A roulette wheel used in a U.S. casino has 38 slots, of which 18 are red, 18 are black, and 2 are green. A roulette wheel used in a French casino has 37 slots, of which 18 are red, 18 are black, and 1 is green. A ball is rolled around the wheel and ends up in one of the slots with equal probability. Suppose that a player bets on red. If a \$1 bet is placed, the player wins \$1 if the ball ends up in a red slot. (The player's \$1 bet is returned.) If the ball ends up in a black or green slot, the player loses \$1. Find the expected value of this game to the player in

(a) The United States.

(b) France.

2.2-10. In the casino game called **high–low**, there are three possible bets. Assume that \$1 is the size of the bet. A pair of fair six-sided dice is rolled and their sum is calculated. If you bet **low**, you win \$1 if the sum of the dice is

{2, 3, 4, 5, 6}. If you bet **high**, you win \$1 if the sum of the dice is {8, 9, 10, 11, 12}. If you bet on {7}, you win \$4 if a sum of 7 is rolled. Otherwise, you lose on each of the three bets. In all three cases, your original dollar is returned if you win. Find the expected value of the game to the bettor for each of these three bets.

2.2-11. In the gambling game craps (see Exercise 1.3-13), the player wins \$1 with probability 0.49293 and loses \$1 with probability 0.50707 for each \$1 bet. What is the expected value of the game to the player?

2.2-12. Suppose that a school has 20 classes: 16 with 25 students in each, three with 100 students in each, and one with 300 students, for a total of 1000 students.

(a) What is the average class size?

(b) Select a student randomly out of the 1000 students. Let the random variable X equal the size of the class to which this student belongs, and define the pmf of X.

(c) Find $E(X)$, the expected value of X. Does this answer surprise you?

2.3 SPECIAL MATHEMATICAL EXPECTATIONS

Let us consider an example in which $x \in \{1, 2, 3\}$ and the pmf is given by $f(1) = 3/6, f(2) = 2/6, f(3) = 1/6$. That is, the probability that the random variable X equals 1, denoted by $P(X = 1)$, is $f(1) = 3/6$. Likewise, $P(X = 2) = f(2) = 2/6$ and $P(X = 3) = f(3) = 1/6$. Of course, $f(x) > 0$ when $x \in S$, and it must be the case that

$$\sum_{x \in S} f(x) = f(1) + f(2) + f(3) = 1.$$

We can think of the points 1, 2, 3 as having weights (probabilities) 3/6, 2/6, 1/6, and their weighted mean (weighted average) is

$$\mu = E(X) = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6} = \frac{5}{3},$$

which, in this illustration, does not equal one of the x values in S. As a matter of fact, it is two thirds of the way between $x = 1$ and $x = 2$.

In Section 2.2 we called $\mu = E(X)$ the mean of the random variable X (or of its distribution). In general, suppose the random variable X has the space $S = \{u_1, u_2, \ldots, u_k\}$ and these points have respective probabilities $P(X = u_i) = f(u_i) > 0$, where $f(x)$ is the pmf. Of course,

$$\sum_{x \in S} f(x) = 1$$

and the **mean** of the random variable X (or of its distribution) is

$$\mu = \sum_{x \in S} x f(x) = u_1 f(u_1) + u_2 f(u_2) + \cdots + u_k f(u_k).$$

That is, in the notation of Section 2.2, $\mu = E(X)$.

Now, u_i is the distance of that ith point from the origin. In mechanics, the product of a distance and its weight is called a moment, so $u_i f(u_i)$ is a moment having a moment arm of length u_i. The sum of such products would be the moment of the system of distances and weights. Actually, it is called the first moment about the origin, since the distances are simply to the first power and the lengths of the arms (distances) are measured from the origin. However, if we compute the first moment about the mean μ, then, since here a moment arm equals $(x - \mu)$, we have

$$\begin{aligned}\sum_{x \in S} (x - \mu) f(x) = E[(X - \mu)] &= E(X) - E(\mu) \\ &= \mu - \mu = 0.\end{aligned}$$

That is, that first moment about μ is equal to zero. In mechanics μ is called the centroid. The last equation implies that if a fulcrum is placed at the centroid μ, then the system of weights would balance, as the sum of the positive moments (when $x > \mu$) about μ equals the sum of the negative moments (when $x < \mu$). In our first illustration, $\mu = 10/6$ is the centroid, so the negative moment

$$\left(1 - \frac{10}{6}\right) \cdot \frac{3}{6} = -\frac{12}{36} = -\frac{1}{3}$$

equals the sum of the two positive moments

$$\left(2 - \frac{10}{6}\right) \cdot \frac{2}{6} + \left(3 - \frac{10}{6}\right) \cdot \frac{1}{6} = \frac{12}{36} = \frac{1}{3}.$$

Since $\mu = E(X)$, it follows from Example 2.2-4 that $b = \mu$ minimizes $E[(X-b)^2]$. Also, Example 2.2-5 shows that

$$\mu = n\left(\frac{N_1}{N}\right)$$

is the mean of the hypergeometric distribution. Moreover, $\mu = 1/p$ is the mean of the geometric distribution from Example 2.2-6.

Statisticians often find it valuable to compute the second moment about the mean μ. It is called the second moment because the distances are raised to the second power, and it is equal to $E[(X - \mu)^2]$; that is,

$$\sum_{x \in S} (x - \mu)^2 f(x) = (u_1 - \mu)^2 f(u_1) + (u_2 - \mu)^2 f(u_2) + \cdots + (u_k - \mu)^2 f(u_k).$$

This weighted mean of the squares of those distances is called the **variance** of the random variable X (or of its distribution). The positive square root of the variance is called the **standard deviation** of X and is denoted by the Greek letter σ (sigma). Thus, the variance is σ^2, sometimes denoted by $\mathrm{Var}(X)$. That is, $\sigma^2 = E[(X - \mu)^2] = \mathrm{Var}(X)$. In our first illustration, since $\mu = 10/6$, the variance equals

$$\sigma^2 = \mathrm{Var}(X) = \left(1 - \frac{10}{6}\right)^2 \cdot \frac{3}{6} + \left(2 - \frac{10}{6}\right)^2 \cdot \frac{2}{6} + \left(3 - \frac{10}{6}\right)^2 \cdot \frac{1}{6} = \frac{120}{216} = \frac{5}{9}.$$

Hence, the standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{120}{216}} = 0.745.$$

It is worth noting that the variance can be computed in another way, because

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

That is, the variance σ^2 equals the difference of the second moment about the origin and the square of the mean. For our first illustration,

$$\begin{aligned} \sigma^2 &= \sum_{x=1}^{3} x^2 f(x) - \mu^2 \\ &= 1^2\left(\frac{3}{6}\right) + 2^2\left(\frac{2}{6}\right) + 3^2\left(\frac{1}{6}\right) - \left(\frac{10}{6}\right)^2 = \frac{20}{6} - \frac{100}{36} = \frac{120}{216} = \frac{5}{9}, \end{aligned}$$

which agrees with our previous computation.

Example 2.3-1 Let X equal the number of spots on the side facing upward after a fair six-sided die is rolled. A reasonable probability model is given by the pmf

$$f(x) = P(X = x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6.$$

The mean of X is

$$\mu = E(X) = \sum_{x=1}^{6} x\left(\frac{1}{6}\right) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}.$$

The second moment about the origin is

$$E(X^2) = \sum_{x=1}^{6} x^2\left(\frac{1}{6}\right) = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}.$$

Thus, the variance equals

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182-147}{12} = \frac{35}{12}.$$

The standard deviation is $\sigma = \sqrt{35/12} = 1.708$. ■

Although most students understand that $\mu = E(X)$ is, in some sense, a measure of the middle of the distribution of X, it is more difficult to get much of a feeling for the variance and the standard deviation. The next example illustrates that the standard deviation is a measure of the dispersion, or spread, of the points belonging to the space S.

Example 2.3-2 Let X have the pmf $f(x) = 1/3$, $x = -1, 0, 1$. Here the mean is

$$\mu = \sum_{x=-1}^{1} xf(x) = (-1)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) = 0.$$

Accordingly, the variance, denoted by σ_X^2, is

$$\begin{aligned}
\sigma_X^2 &= E[(X-0)^2] \\
&= \sum_{x=-1}^{1} x^2 f(x) \\
&= (-1)^2\left(\frac{1}{3}\right) + (0)^2\left(\frac{1}{3}\right) + (1)^2\left(\frac{1}{3}\right) \\
&= \frac{2}{3},
\end{aligned}$$

so the standard deviation is $\sigma_X = \sqrt{2/3}$. Next, let another random variable Y have the pmf $g(y) = 1/3$, $y = -2, 0, 2$. Its mean is also zero, and it is easy to show that $\mathrm{Var}(Y) = 8/3$, so the standard deviation of Y is $\sigma_Y = 2\sqrt{2/3}$. Here the standard deviation of Y is twice that of the standard deviation of X, reflecting the fact that the probability of Y is spread out twice as much as that of X. ■

Example 2.3-3

Let X have a uniform distribution on the first m positive integers. The mean of X is

$$\mu = E(X) = \sum_{x=1}^{m} x\left(\frac{1}{m}\right) = \frac{1}{m}\sum_{x=1}^{m} x$$
$$= \left(\frac{1}{m}\right)\frac{m(m+1)}{2} = \frac{m+1}{2}.$$

To find the variance of X, we first find

$$E(X^2) = \sum_{x=1}^{m} x^2\left(\frac{1}{m}\right) = \frac{1}{m}\sum_{x=1}^{m} x^2$$
$$= \left(\frac{1}{m}\right)\frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}.$$

Thus, the variance of X is

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2]$$
$$= E(X^2) - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2$$
$$= \frac{m^2-1}{12}.$$

For example, we find that if X equals the outcome when rolling a fair six-sided die, the pmf of X is

$$f(x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6;$$

the respective mean and variance of X are

$$\mu = \frac{6+1}{2} = 3.5 \qquad \text{and} \qquad \sigma^2 = \frac{6^2-1}{12} = \frac{35}{12},$$

which agrees with calculations of Example 2.3-1. ■

Now let X be a random variable with mean μ_X and variance σ_X^2. Of course, $Y = aX + b$, where a and b are constants, is a random variable, too. The mean of Y is

$$\mu_Y = E(Y) = E(aX+b) = aE(X) + b = a\mu_X + b.$$

Moreover, the variance of Y is

$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = E[(aX+b-a\mu_X-b)^2] = E[a^2(X-\mu_X)^2] = a^2\sigma_X^2.$$

Thus, $\sigma_Y = |a|\sigma_X$. To illustrate, note in Example 2.3-2 that the relationship between the two distributions could be explained by defining $Y = 2X$, so that $\sigma_Y^2 = 4\sigma_X^2$ and consequently $\sigma_Y = 2\sigma_X$, which we had observed there. In addition, we see that adding or subtracting a constant from X does not change the variance. For illustration, $\mathrm{Var}(X-1) = \mathrm{Var}(X)$, because $a = 1$ and $b = -1$. Also note that $\mathrm{Var}(-X) = \mathrm{Var}(X)$ because here $a = -1$ and $b = 0$.

Let r be a positive integer. If

$$E(X^r) = \sum_{x \in S} x^r f(x)$$

is finite, it is called the rth **moment** of the distribution about the origin. In addition, the expectation

$$E[(X-b)^r] = \sum_{x \in S} (x-b)^r f(x)$$

is called the rth moment of the distribution about b.

For a given positive integer r,

$$E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)]$$

is called the rth factorial moment. We note that the second factorial moment is equal to the difference of the second and first moments about 0:

$$E[X(X-1)] = E(X^2) - E(X).$$

There is another formula that can be used to compute the variance. This formula uses the second factorial moment and sometimes simplifies the calculations. First find the values of $E(X)$ and $E[X(X-1)]$. Then

$$\sigma^2 = E[X(X-1)] + E(X) - [E(X)]^2,$$

since, by the distributive property of E, this becomes

$$\sigma^2 = E(X^2) - E(X) + E(X) - [E(X)]^2 = E(X^2) - \mu^2.$$

Example 2.3-4 In Example 2.2-5 concerning the hypergeometric distribution, we found that the mean of that distribution is

$$\mu = E(X) = n\left(\frac{N_1}{N}\right) = np,$$

where $p = N_1/N$, the fraction of red chips in the N chips. In Exercise 2.3-10, it is determined that

$$E[X(X-1)] = \frac{(n)(n-1)(N_1)(N_1-1)}{N(N-1)}.$$

Thus, the variance of X is $E[X(X-1)] + E(X) - [E(X)]^2$, namely,

$$\sigma^2 = \frac{n(n-1)(N_1)(N_1-1)}{N(N-1)} + \frac{nN_1}{N} - \left(\frac{nN_1}{N}\right)^2.$$

After some straightforward algebra, we find that

$$\sigma^2 = n\left(\frac{N_1}{N}\right)\left(\frac{N_2}{N}\right)\left(\frac{N-n}{N-1}\right) = np(1-p)\left(\frac{N-n}{N-1}\right). \quad \blacksquare$$

We now define a function that will help us generate the moments of a distribution. Thus, this function is called the moment-generating function. Although this generating characteristic is extremely important, there is a uniqueness property that is even more important. We first define the new function and then explain this uniqueness property before showing how it can be used to compute the moments of X.

Definition 2.3-1
Let X be a random variable of the discrete type with pmf $f(x)$ and space S. If there is a positive number h such that

$$E(e^{tX}) = \sum_{x \in S} e^{tx} f(x)$$

exists and is finite for $-h < t < h$, then the function defined by

$$M(t) = E(e^{tX})$$

is called the **moment-generating function of** X (or of the distribution of X). This function is often abbreviated as mgf.

First, it is evident that if we set $t = 0$, we have $M(0) = 1$. Moreover, if the space of S is $\{b_1, b_2, b_3, \ldots\}$, then the moment-generating function is given by the expansion

$$M(t) = e^{tb_1} f(b_1) + e^{tb_2} f(b_2) + e^{tb_3} f(b_3) + \cdots .$$

Thus, the coefficient of e^{tb_i} is the probability

$$f(b_i) = P(X = b_i).$$

Accordingly, if two random variables (or two distributions of probability) have the same moment-generating function, they must have the same distribution of probability. That is, if the two random variables had the two probability mass functions $f(x)$ and $g(y)$, as well as the same space $S = \{b_1, b_2, b_3, \ldots\}$, and if

$$e^{tb_1} f(b_1) + e^{tb_2} f(b_2) + \cdots = e^{tb_1} g(b_1) + e^{tb_2} g(b_2) + \cdots \qquad (2.3\text{-}1)$$

for all t, $-h < t < h$, then mathematical transform theory requires that

$$f(b_i) = g(b_i), \qquad i = 1, 2, 3, \ldots .$$

So we see that the moment-generating function of a discrete random variable uniquely determines the distribution of that random variable. In other words, if the mgf exists, there is one and only one distribution of probability associated with that mgf.

REMARK From elementary algebra, we can get some understanding of why Equation 2.3-1 requires that $f(b_i) = g(b_i)$. In that equation, let $e^t = w$ and say the points in the support, namely, $b_1, b_2, \ldots, b_k$, are positive integers, the largest of which is m. Then Equation 2.3-1 provides the equality of two mth-degree polynomials in w for an uncountable number of values of w. A fundamental theorem of algebra requires that the corresponding coefficients of the two polynomials be equal; that is, $f(b_i) = g(b_i)$, $i = 1, 2, \ldots, k$. ■

Example 2.3-5 If X has the mgf

$$M(t) = e^t \left(\frac{3}{6}\right) + e^{2t} \left(\frac{2}{6}\right) + e^{3t} \left(\frac{1}{6}\right), \qquad -\infty < t < \infty,$$

then the support of X is $S = \{1, 2, 3\}$ and the associated probabilities are

$$P(X = 1) = \frac{3}{6}, \qquad P(X = 2) = \frac{2}{6}, \qquad P(X = 3) = \frac{1}{6}.$$

We could write this, if we choose to do so, by saying that X has the pmf

$$f(x) = \frac{4-x}{6}, \qquad x = 1, 2, 3.$$

Example 2.3-6 Suppose the mgf of X is

$$M(t) = \frac{e^t/2}{1 - e^t/2}, \qquad t < \ln 2.$$

Until we expand $M(t)$, we cannot detect the coefficients of $e^{b_i t}$. Recalling that

$$(1-z)^{-1} = 1 + z + z^2 + z^3 + \cdots, \qquad -1 < z < 1,$$

we have

$$\begin{aligned} \frac{e^t}{2}\left(1 - \frac{e^t}{2}\right)^{-1} &= \frac{e^t}{2}\left(1 + \frac{e^t}{2} + \frac{e^{2t}}{2^2} + \frac{e^{3t}}{2^3} + \cdots\right) \\ &= (e^t)\left(\frac{1}{2}\right)^1 + \left(e^{2t}\right)\left(\frac{1}{2}\right)^2 + \left(e^{3t}\right)\left(\frac{1}{2}\right)^3 + \cdots \end{aligned}$$

when $e^t/2 < 1$ and thus $t < \ln 2$. That is,

$$P(X = x) = \left(\frac{1}{2}\right)^x$$

when x is a positive integer, or, equivalently, the pmf of X is

$$f(x) = \left(\frac{1}{2}\right)^x, \qquad x = 1, 2, 3, \ldots.$$

From the theory of Laplace transforms, it can be shown that the existence of $M(t)$, for $-h < t < h$, implies that derivatives of $M(t)$ of all orders exist at $t = 0$; hence, $M(t)$ is continuous at $t = 0$. Moreover, it is permissible to interchange differentiation and summation as the series converges uniformly. Thus,

$$\begin{aligned} M'(t) &= \sum_{x \in S} x e^{tx} f(x), \\ M''(t) &= \sum_{x \in S} x^2 e^{tx} f(x), \end{aligned}$$

and for each positive integer r,

$$M^{(r)}(t) = \sum_{x \in S} x^r e^{tx} f(x).$$

Setting $t = 0$, we see that

$$\begin{aligned} M'(0) &= \sum_{x \in S} x f(x) \;=\; E(X), \\ M''(0) &= \sum_{x \in S} x^2 f(x) \;=\; E(X^2), \end{aligned}$$

and, in general,

$$M^{(r)}(0) = \sum_{x \in S} x^r f(x) \;=\; E(X^r).$$

In particular, if the moment-generating function exists, then

$$M'(0) = E(X) = \mu \qquad \text{and} \qquad M''(0) - [M'(0)]^2 = E(X^2) - [E(X)]^2 = \sigma^2.$$

The preceding argument shows that we can find the moments of X by differentiating $M(t)$. In using this technique, it must be emphasized that first we evaluate the summation representing $M(t)$ to obtain a closed-form solution and then we differentiate that solution to obtain the moments of X. The next example illustrates the use of the moment-generating function for finding the first and second moments and then the mean and variance of the geometric distribution.

Example 2.3-7

Suppose X has the geometric distribution of Example 2.2-6; that is, the pmf of X is

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

Then the mgf of X is

$$\begin{aligned} M(t) = E(e^{tX}) \;&=\; \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = \left(\frac{p}{q}\right) \sum_{x=1}^{\infty} (qe^t)^x \\ &= \left(\frac{p}{q}\right) [(qe^t) + (qe^t)^2 + (qe^t)^3 + \cdots] \\ &= \left(\frac{p}{q}\right) \frac{qe^t}{1 - qe^t} = \frac{pe^t}{1 - qe^t}, \qquad \text{provided } qe^t < 1 \text{ or } t < -\ln q. \end{aligned}$$

Note that $-\ln q = h$ is positive. To find the mean and the variance of X, we first differentiate $M(t)$ twice:

$$M'(t) = \frac{(1 - qe^t)(pe^t) - pe^t(-qe^t)}{(1 - qe^t)^2} = \frac{pe^t}{(1 - qe^t)^2}$$

and

$$M''(t) = \frac{(1 - qe^t)^2 pe^t - pe^t(2)(1 - qe^t)(-qe^t)}{(1 - qe^t)^4} = \frac{pe^t(1 + qe^t)}{(1 - qe^t)^3}.$$

Of course, $M(0) = 1$ and $M(t)$ is continuous at $t = 0$ as we were able to differentiate at $t = 0$. With $1 - q = p$,

$$M'(0) = \frac{p}{(1 - q)^2} = \frac{1}{p} = \mu$$

and

$$M''(0) = \frac{p(1 + q)}{(1 - q)^3} = \frac{1 + q}{p^2}.$$

Thus,

$$\sigma^2 = M''(0) - [M'(0)]^2 = \frac{1 + q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Exercises

2.3-1. Find the mean and variance for the following discrete distributions:

(a) $f(x) = \dfrac{1}{5}, \qquad x = 5, 10, 15, 20, 25.$

(b) $f(x) = 1, \qquad x = 5.$

(c) $f(x) = \dfrac{4-x}{6}, \qquad x = 1, 2, 3.$

2.3-2. For each of the following distributions, find $\mu = E(X)$, $E[X(X-1)]$, and $\sigma^2 = E[X(X-1)] + E(X) - \mu^2$:

(a) $f(x) = \dfrac{3!}{x!(3-x)!}\left(\dfrac{1}{4}\right)^x\left(\dfrac{3}{4}\right)^{3-x}, \qquad x = 0, 1, 2, 3.$

(b) $f(x) = \dfrac{4!}{x!(4-x)!}\left(\dfrac{1}{2}\right)^4, \qquad x = 0, 1, 2, 3, 4.$

2.3-3. Given $E(X+4) = 10$ and $E[(X+4)^2] = 116$, determine **(a)** $\text{Var}(X+4)$, **(b)** $\mu = E(X)$, and **(c)** $\sigma^2 = \text{Var}(X)$.

2.3-4. Let μ and σ^2 denote the mean and variance of the random variable X. Determine $E[(X-\mu)/\sigma]$ and $E\{[(X-\mu)/\sigma]^2\}$.

2.3-5. Consider an experiment that consists of selecting a card at random from an ordinary deck of cards. Let the random variable X equal the value of the selected card, where Ace = 1, Jack = 11, Queen = 12, and King = 13. Thus, the space of X is $S = \{1, 2, 3, \ldots, 13\}$. If the experiment is performed in an unbiased manner, assign probabilities to these 13 outcomes and compute the mean μ of this probability distribution.

2.3-6. Place eight chips in a bowl: Three have the number 1 on them, two have the number 2, and three have the number 3. Say each chip has a probability of 1/8 of being drawn at random. Let the random variable X equal the number on the chip that is selected, so that the space of X is $S = \{1, 2, 3\}$. Make reasonable probability assignments to each of these three outcomes, and compute the mean μ and the variance σ^2 of this probability distribution.

2.3-7. Let X equal an integer selected at random from the first m positive integers, $\{1, 2, \ldots, m\}$. Find the value of m for which $E(X) = \text{Var}(X)$. (See Zerger in the references.)

2.3-8. Let X equal the larger outcome when a pair of fair four-sided dice is rolled. The pmf of X is

$$f(x) = \frac{2x-1}{16}, \qquad x = 1, 2, 3, 4.$$

Find the mean, variance, and standard deviation of X.

2.3-9. A warranty is written on a product worth \$10,000 so that the buyer is given \$8000 if it fails in the first year, \$6000 if it fails in the second, \$4000 if it fails in the third, \$2000 if it fails in the fourth, and zero after that. The probability that the product fails in the first year is 0.1, and the probability that it fails in any subsequent year, provided that it did not fail prior to that year, is 0.1. What is the expected value of the warranty?

2.3-10. To find the variance of a hypergeometric random variable in Example 2.3-4 we used the fact that

$$E[X(X-1)] = \frac{N_1(N_1-1)(n)(n-1)}{N(N-1)}.$$

Prove this result by making the change of variables $k = x - 2$ and noting that

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

2.3-11. If the moment-generating function of X is

$$M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t},$$

find the mean, variance, and pmf of X.

2.3-12. Let X equal the number of people selected at random that you must ask in order to find someone with the same birthday as yours. Assume that each day of the year is equally likely, and ignore February 29.

(a) What is the pmf of X?

(b) Give the values of the mean, variance, and standard deviation of X.

(c) Find $P(X > 400)$ and $P(X < 300)$.

2.3-13. For each question on a multiple-choice test, there are five possible answers, of which exactly one is correct. If a student selects answers at random, give the probability that the first question answered correctly is question 4.

2.3-14. The probability that a machine produces a defective item is 0.01. Each item is checked as it is produced. Assume that these are independent trials, and compute the probability that at least 100 items must be checked to find one that is defective.

2.3-15. Apples are packaged automatically in 3-pound bags. Suppose that 4% of the time the bag of apples weighs less than 3 pounds. If you select bags randomly and weigh them in order to discover one underweight bag of apples, find the probability that the number of bags that must be selected is

(a) At least 20.

(b) At most 20.

(c) Exactly 20.

2.3-16. Let X equal the number of flips of a fair coin that are required to observe the same face on consecutive flips.

(a) Find the pmf of X. HINT: Draw a tree diagram.

(b) Find the moment-generating function of X.

(c) Use the mgf to find the values of **(i)** the mean and **(ii)** the variance of X.

(d) Find the values of **(i)** $P(X \leq 3)$, **(ii)** $P(X \geq 5)$, and **(iii)** $P(X = 3)$.

2.3-17. Let X equal the number of flips of a fair coin that are required to observe heads–tails on consecutive flips.

(a) Find the pmf of X. HINT: Draw a tree diagram.

(b) Show that the mgf of X is $M(t) = e^{2t}/(e^t - 2)^2$.

(c) Use the mgf to find the values of **(i)** the mean and **(ii)** the variance of X.

(d) Find the values of **(i)** $P(X \leq 3)$, **(ii)** $P(X \geq 5)$, and **(iii)** $P(X = 3)$.

2.3-18. Let X have a geometric distribution. Show that

$$P(X > k + j \mid X > k) = P(X > j),$$

where k and j are nonnegative integers. NOTE: We sometimes say that in this situation there has been loss of memory.

2.3-19. Given a random permutation of the integers in the set $\{1, 2, 3, 4, 5\}$, let X equal the number of integers that are in their natural position. The moment-generating function of X is

$$M(t) = \frac{44}{120} + \frac{45}{120}e^t + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}.$$

(a) Find the mean and variance of X.

(b) Find the probability that at least one integer is in its natural position.

(c) Draw a graph of the probability histogram of the pmf of X.

2.4 THE BINOMIAL DISTRIBUTION

The probability models for random experiments that will be described in this section occur frequently in applications.

A **Bernoulli experiment** is a random experiment, the outcome of which can be classified in one of two mutually exclusive and exhaustive ways—say, success or failure (e.g., female or male, life or death, nondefective or defective). A sequence of **Bernoulli trials** occurs when a Bernoulli experiment is performed several *independent* times and the probability of success—say, p—remains the *same* from trial to trial. That is, in such a sequence we let p denote the probability of success on each trial. In addition, we shall frequently let $q = 1 - p$ denote the probability of failure; that is, we shall use q and $1 - p$ interchangeably.

Example 2.4-1 Suppose that the probability of germination of a beet seed is 0.8 and the germination of a seed is called a success. If we plant 10 seeds and can assume that the germination of one seed is independent of the germination of another seed, this would correspond to 10 Bernoulli trials with $p = 0.8$. ■

Example 2.4-2 In the Michigan daily lottery the probability of winning when placing a six-way boxed bet is 0.006. A bet placed on each of 12 successive days would correspond to 12 Bernoulli trials with $p = 0.006$. ■

Let X be a random variable associated with a Bernoulli trial by defining it as follows:

$$X\,(\text{success}) = 1 \qquad \text{and} \qquad X\,(\text{failure}) = 0.$$

That is, the two outcomes, success and failure, are denoted by one and zero, respectively. The pmf of X can be written as

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

and we say that X has a **Bernoulli distribution**. The expected value of X is

$$\mu = E(X) = \sum_{x=0}^{1} x\,p^x(1-p)^{1-x} = (0)(1-p) + (1)(p) = p,$$

and the variance of X is

$$\begin{aligned}\sigma^2 &= \operatorname{Var}(X) = \sum_{x=0}^{1}(x-p)^2p^x(1-p)^{1-x}\\ &= (0-p)^2(1-p) + (1-p)^2p = p(1-p) = pq.\end{aligned}$$

It follows that the standard deviation of X is

$$\sigma = \sqrt{p(1-p)} = \sqrt{pq}.$$

In a sequence of n Bernoulli trials, we shall let X_i denote the Bernoulli random variable associated with the ith trial. An observed sequence of n Bernoulli trials will then be an n-tuple of zeros and ones, and we often call this collection a **random sample** of size n from a Bernoulli distribution.

Example 2.4-3

Out of millions of instant lottery tickets, suppose that 20% are winners. If five such tickets are purchased, then $(0, 0, 0, 1, 0)$ is a possible observed sequence in which the fourth ticket is a winner and the other four are losers. Assuming independence among winning and losing tickets, we observe that the probability of this outcome is

$$(0.8)(0.8)(0.8)(0.2)(0.8) = (0.2)(0.8)^4.$$ ■

Example 2.4-4

If five beet seeds are planted in a row, a possible observed sequence would be (1, 0, 1, 0, 1) in which the first, third, and fifth seeds germinated and the other two did not. If the probability of germination is $p = 0.8$, the probability of this outcome is, assuming independence,

$$(0.8)(0.2)(0.8)(0.2)(0.8) = (0.8)^3(0.2)^2.$$ ■

In a sequence of Bernoulli trials, we are often interested in the total number of successes but not the actual order of their occurrences. If we let the random variable X equal the number of observed successes in n Bernoulli trials, then the possible values of X are $0, 1, 2, \ldots, n$. If x successes occur, where $x = 0, 1, 2, \ldots, n$, then $n - x$ failures occur. The number of ways of selecting x positions for the x successes in the n trials is

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

Since the trials are independent and since the probabilities of success and failure on each trial are, respectively, p and $q = 1 - p$, the probability of each of these ways is $p^x(1-p)^{n-x}$. Thus, $f(x)$, the pmf of X, is the sum of the probabilities of the $\binom{n}{x}$ mutually exclusive events; that is,

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

These probabilities are called binomial probabilities, and the random variable X is said to have a **binomial distribution**.

Summarizing, a binomial experiment satisfies the following properties:

1. A Bernoulli (success–failure) experiment is performed n times, where n is a (non-random) constant.
2. The trials are independent.
3. The probability of success on each trial is a constant p; the probability of failure is $q = 1 - p$.
4. The random variable X equals the number of successes in the n trials.

A binomial distribution will be denoted by the symbol $b(n, p)$, and we say that the distribution of X is $b(n, p)$. The constants n and p are called the **parameters** of the binomial distribution; they correspond to the number n of independent trials and the probability p of success on each trial. Thus, if we say that the distribution of X is $b(12, 1/4)$, we mean that X is the number of successes in a random sample of size $n = 12$ from a Bernoulli distribution with $p = 1/4$.

Example 2.4-5

In the instant lottery with 20% winning tickets, if X is equal to the number of winning tickets among $n = 8$ that are purchased, then the probability of purchasing two winning tickets is

$$f(2) = P(X = 2) = \binom{8}{2} (0.2)^2 (0.8)^6 = 0.2936.$$

The distribution of the random variable X is $b(8, 0.2)$. ■

Example 2.4-6

In order to obtain a better feeling for the effect of the parameters n and p on the distribution of probabilities, four probability histograms are displayed in Figure 2.4-1. ■

Example 2.4-7

In Example 2.4-1, the number X of seeds that germinate in $n = 10$ independent trials is $b(10, 0.8)$; that is,

$$f(x) = \binom{10}{x} (0.8)^x (0.2)^{10-x}, \qquad x = 0, 1, 2, \ldots, 10.$$

In particular,

$$\begin{aligned} P(X \le 8) &= 1 - P(X = 9) - P(X = 10) \\ &= 1 - 10(0.8)^9(0.2) - (0.8)^{10} \;=\; 0.6242. \end{aligned}$$

Also, with a little more work, we could compute

$$P(X \le 6) = \sum_{x=0}^{6} \binom{10}{x} (0.8)^x (0.2)^{10-x} = 0.1209.$$ ■

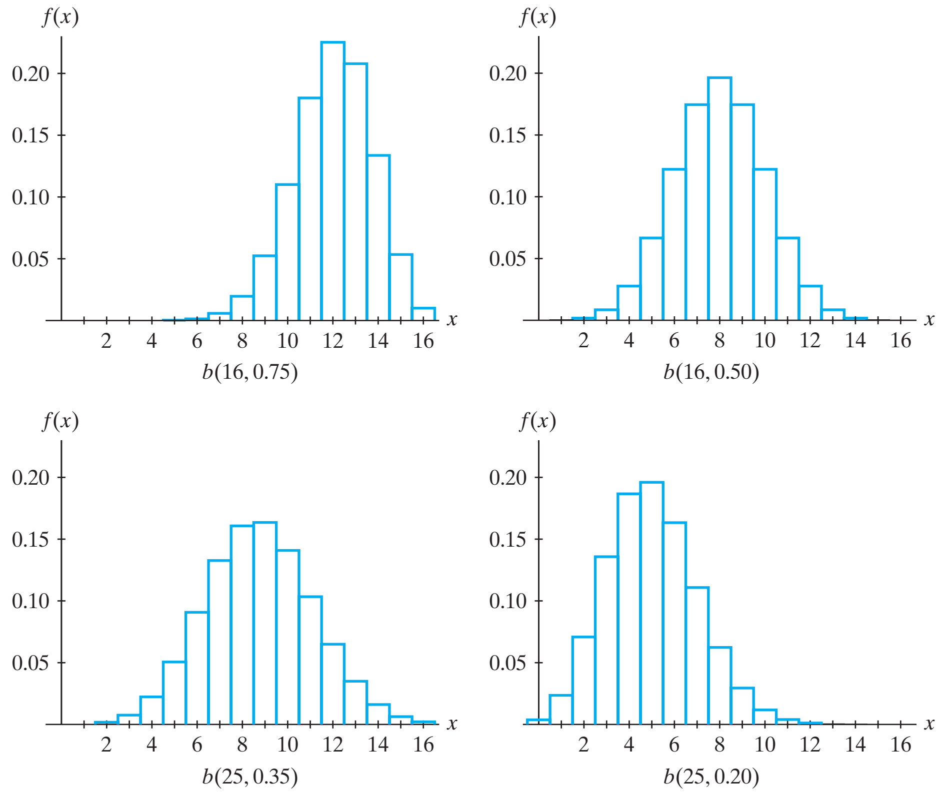

Figure 2.4-1 Binomial probability histograms

Recall that cumulative probabilities like those in the previous example are given by the cumulative distribution function (cdf) of X or sometimes called the distribution function (df) of X, defined by

$$F(x) = P(X \le x), \qquad -\infty < x < \infty.$$

We tend to use the cdf (rather than the pmf) to obtain probabilities of events involving a $b(n, p)$ random variable X. Tables of this cdf are given in Table II in Appendix B for selected values of n and p.

For the binomial distribution given in Example 2.4-7, namely, the $b(10, 0.8)$ distribution, the distribution function is defined by

$$F(x) = P(X \le x) = \sum_{y=0}^{\lfloor x \rfloor} \binom{10}{y} (0.8)^y (0.2)^{10-y},$$

where $\lfloor x \rfloor$ is the greatest integer in x. A graph of this cdf is shown in Figure 2.4-2. Note that the vertical jumps at the integers in this step function are equal to the probabilities associated with those respective integers.

Example 2.4-8 Leghorn chickens are raised for laying eggs. Let $p = 0.5$ be the probability that a newly hatched chick is a female. Assuming independence, let X equal the number of female chicks out of 10 newly hatched chicks selected at random. Then the distribution of X is $b(10, 0.5)$. From Table II in Appendix B, the probability of 5 or fewer female chicks is

$$P(X \le 5) = 0.6230.$$

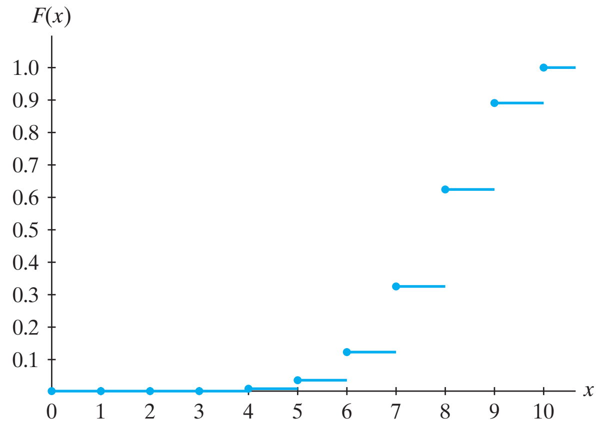

Figure 2.4-2 Distribution function for the $b(10, 0.8)$ distribution

The probability of exactly 6 female chicks is

$$P(X = 6) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4$$
$$= P(X \le 6) - P(X \le 5)$$
$$= 0.8281 - 0.6230 \;=\; 0.2051,$$

since $P(X \le 6) = 0.8281$. The probability of at least 6 female chicks is

$$P(X \ge 6) = 1 - P(X \le 5) = 1 - 0.6230 = 0.3770.$$

Although probabilities for the binomial distribution $b(n,p)$ are given in Table II in Appendix B for selected values of p that are less than or equal to 0.5, the next example demonstrates that this table can also be used for values of p that are greater than 0.5. In later sections we learn how to approximate certain binomial probabilities with those of other distributions. In addition, you may use your calculator and/or a statistical package such as Minitab to find binomial probabilities.

Example 2.4-9 Suppose that we are in one of those rare times when 65% of the American public approve of the way the president of the United States is handling the job. Take a random sample of $n = 8$ Americans and let Y equal the number who give approval. Then, to a very good approximation, the distribution of Y is $b(8, 0.65)$. (Y would have the stated distribution exactly if the sampling were done with replacement, but most public opinion polling uses sampling without replacement.) To find $P(Y \ge 6)$, note that

$$P(Y \ge 6) = P(8 - Y \le 8 - 6) = P(X \le 2),$$

where $X = 8 - Y$ counts the number who disapprove. Since $q = 1 - p = 0.35$ equals the probability of disapproval by each person selected, the distribution of X is $b(8, 0.35)$. (See Figure 2.4-3.) From Table II in Appendix B, since $P(X \le 2) = 0.4278$, it follows that $P(Y \ge 6) = 0.4278$.

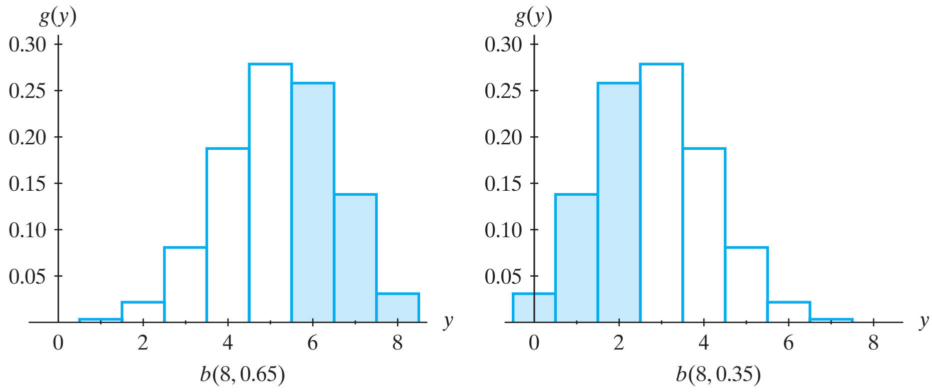

Figure 2.4-3 Presidential approval histogram

Similarly,

$$\begin{aligned} P(Y \le 5) &= P(8 - Y \ge 8 - 5) \\ &= P(X \ge 3) \;=\; 1 - P(X \le 2) \\ &= 1 - 0.4278 \;=\; 0.5722 \end{aligned}$$

and

$$\begin{aligned} P(Y = 5) &= P(8 - Y = 8 - 5) \\ &= P(X = 3) \;=\; P(X \le 3) - P(X \le 2) \\ &= 0.7064 - 0.4278 \;=\; 0.2786. \end{aligned}$$

Recall that if n is a positive integer, then

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

Thus, if we use this binomial expansion with $b = p$ and $a = 1 - p$, then the sum of the binomial probabilities is

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

a result that had to follow from the fact that $f(x)$ is a pmf.

We now use the binomial expansion to find the mgf for a binomial random variable and then the mean and variance.

The mgf is

$$\begin{aligned} M(t) = E(e^{tX}) \;&=\; \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= \left[(1-p) + pe^t\right]^n, \qquad -\infty < t < \infty, \end{aligned}$$

from the expansion of $(a+b)^n$ with $a = 1-p$ and $b = pe^t$. It is interesting to note that here and elsewhere the mgf is usually rather easy to compute if the pmf has a factor involving an exponential, like p^x in the binomial pmf.

The first two derivatives of $M(t)$ are

$$M'(t) = n[(1-p) + pe^t]^{n-1}(pe^t)$$

and

$$M''(t) = n(n-1)[(1-p) + pe^t]^{n-2}(pe^t)^2 + n[(1-p) + pe^t]^{n-1}(pe^t).$$

Thus,

$$\mu = E(X) = M'(0) = np$$

and

$$\begin{aligned}\sigma^2 = E(X^2) - [E(X)]^2 &= M''(0) - [M'(0)]^2 \\ = n(n-1)p^2 + np - (np)^2 &= np(1-p).\end{aligned}$$

Note that when p is the probability of success on each trial, the expected number of successes in n trials is np, a result that agrees with our intuition.

In the special case when $n = 1$, X has a Bernoulli distribution and

$$M(t) = (1-p) + pe^t$$

for all real values of t, $\mu = p$, and $\sigma^2 = p(1-p)$.

Example 2.4-10 Suppose that observation over a long period of time has disclosed that, on the average, 1 out of 10 items produced by a process is defective. Select five items independently from the production line and test them. Let X denote the number of defective items among the $n = 5$ items. Then X is $b(5, 0.1)$. Furthermore,

$$E(X) = 5(0.1) = 0.5, \qquad \mathrm{Var}(X) = 5(0.1)(0.9) = 0.45.$$

For example, the probability of observing at most one defective item is

$$P(X \le 1) = \binom{5}{0}(0.1)^0(0.9)^5 + \binom{5}{1}(0.1)^1(0.9)^4 = 0.9185.$$ ■

Suppose that an urn contains N_1 success balls and N_2 failure balls. Let $p = N_1/(N_1+N_2)$, and let X equal the number of success balls in a random sample of size n that is taken from this urn. If the sampling is done one at a time with replacement, then the distribution of X is $b(n, p)$; if the sampling is done without replacement, then X has a hypergeometric distribution with pmf

$$f(x) = \frac{\dbinom{N_1}{x}\dbinom{N_2}{n-x}}{\dbinom{N_1+N_2}{n}},$$

where x is a nonnegative integer such that $x \le n$, $x \le N_1$, and $n - x \le N_2$. When $N_1 + N_2$ is large and n is relatively small, it makes little difference if the sampling is done with or without replacement. In Figure 2.4-4, the probability histograms are compared for different combinations of n, N_1, and N_2.

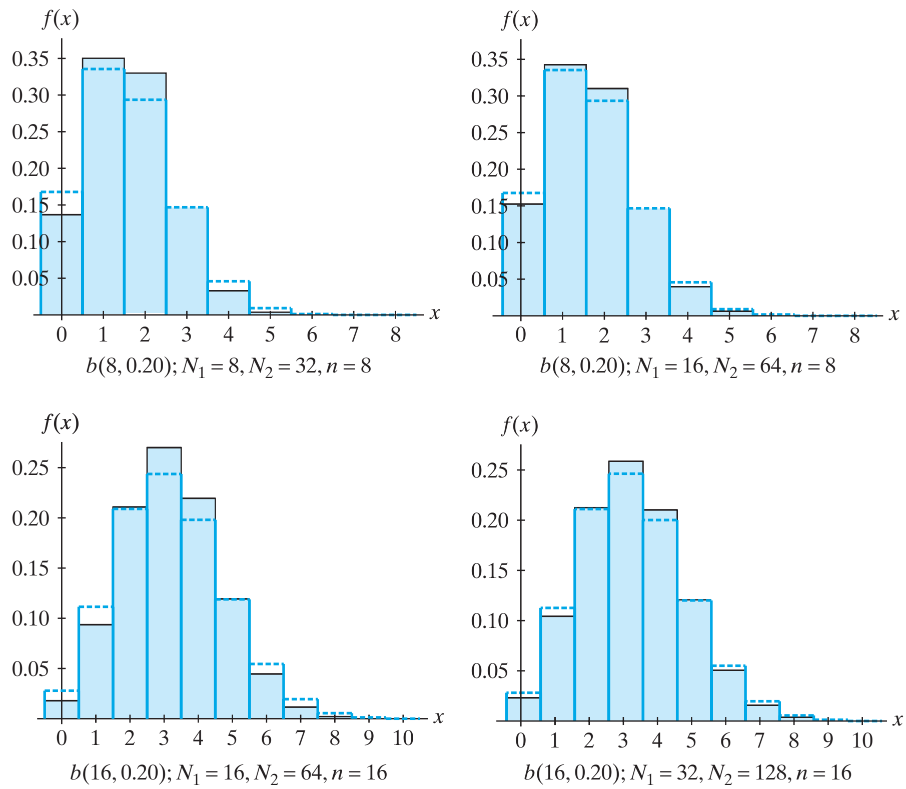

Figure 2.4-4 Binomial and hypergeometric (shaded) probability histograms

Exercises

2.4-1. An urn contains 7 red and 11 white balls. Draw one ball at random from the urn. Let $X = 1$ if a red ball is drawn, and let $X = 0$ if a white ball is drawn. Give the pmf, mean, and variance of X.

2.4-2. Suppose that in Exercise 2.4-1, $X = 1$ if a red ball is drawn and $X = -1$ if a white ball is drawn. Give the pmf, mean, and variance of X.

2.4-3. On a six-question multiple-choice test there are five possible answers for each question, of which one is correct (C) and four are incorrect (I). If a student guesses randomly and independently, find the probability of

(a) Being correct only on questions 1 and 4 (i.e., scoring C, I, I, C, I, I).

(b) Being correct on two questions.

2.4-4. It is claimed that 15% of the ducks in a particular region have patent schistosome infection. Suppose that seven ducks are selected at random. Let X equal the number of ducks that are infected.

(a) Assuming independence, how is X distributed?

(b) Find **(i)** $P(X \geq 2)$, **(ii)** $P(X = 1)$, and **(iii)** $P(X \leq 3)$.

2.4-5. In a lab experiment involving inorganic syntheses of molecular precursors to organometallic ceramics, the final step of a five-step reaction involves the formation of a metal–metal bond. The probability of such a bond forming is $p = 0.20$. Let X equal the number of successful reactions out of $n = 25$ such experiments.

(a) Find the probability that X is at most 4.

(b) Find the probability that X is at least 5.

(c) Find the probability that X is equal to 6.

(d) Give the mean, variance, and standard deviation of X.

2.4-6. It is believed that approximately 75% of American youth now have insurance due to the health care law. Suppose this is true, and let X equal the number of American youth in a random sample of $n = 15$ with private health insurance.

(a) How is X distributed?

(b) Find the probability that X is at least 10.

(c) Find the probability that X is at most 10.

(d) Find the probability that X is equal to 10.

(e) Give the mean, variance, and standard deviation of X.

2.4-7. Suppose that 2000 points are selected independently and at random from the unit square $\{(x, y) : 0 \leq x < 1, 0 \leq y < 1\}$. Let W equal the number of points that fall into $A = \{(x, y) : x^2 + y^2 < 1\}$.

(a) How is W distributed?

(b) Give the mean, variance, and standard deviation of W.

(c) What is the expected value of $W/500$?

(d) Use the computer to select 2000 pairs of random numbers. Determine the value of W and use that value to find an estimate for π. (Of course, we know the real value of π, and more will be said about estimation later in this text.)

(e) How could you extend part (d) to estimate the volume $V = (4/3)\pi$ of a ball of radius 1 in 3-space?

(f) How could you extend these techniques to estimate the "volume" of a ball of radius 1 in n-space?

2.4-8. A boiler has four relief valves. The probability that each opens properly is 0.99.

(a) Find the probability that at least one opens properly.

(b) Find the probability that all four open properly.

2.4-9. Suppose that the percentage of American drivers who are multitaskers (e.g., talk on cell phones, eat a snack, or text message at the same time they are driving) is approximately 80%. In a random sample of $n = 20$ drivers, let X equal the number of multitaskers.

(a) How is X distributed?

(b) Give the values of the mean, variance, and standard deviation of X.

(c) Find the following probabilities: **(i)** $P(X = 15)$, **(ii)** $P(X > 15)$, and **(iii)** $P(X \le 15)$.

2.4-10. A certain type of mint has a label weight of 20.4 grams. Suppose that the probability is 0.90 that a mint weighs more than 20.7 grams. Let X equal the number of mints that weigh more than 20.7 grams in a sample of eight mints selected at random.

(a) How is X distributed if we assume independence?

(b) Find **(i)** $P(X = 8)$, **(ii)** $P(X \le 6)$, and **(iii)** $P(X \ge 6)$.

2.4-11. A random variable X has a binomial distribution with mean 6 and variance 3.6. Find $P(X = 4)$.

2.4-12. In the casino game chuck-a-luck, three fair six-sided dice are rolled. One possible bet is \$1 on fives, and the payoff is equal to \$1 for each five on that roll. In addition, the dollar bet is returned if at least one five is rolled. The dollar that was bet is lost only if no fives are rolled. Let X denote the payoff for this game. Then X can equal -1, 1, 2, or 3.

(a) Determine the pmf $f(x)$.

(b) Calculate μ, σ^2, and σ.

(c) Depict the pmf as a probability histogram.

2.4-13. It is claimed that for a particular lottery, 1/10 of the 50 million tickets will win a prize. What is the probability of winning at least one prize if you purchase **(a)** 10 tickets or **(b)** 15 tickets?

2.4-14. For the lottery described in Exercise 2.4-13, find the smallest number of tickets that must be purchased so that the probability of winning at least one prize is greater than **(a)** 0.50; **(b)** 0.95.

2.4-15. A hospital obtains 40% of its flu vaccine from Company A, 50% from Company B, and 10% from Company C. From past experience, it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective. The hospital tests five vials from each shipment. If at least one of the five is ineffective, find the conditional probability of that shipment's having come from C.

2.4-16. A company starts a fund of M dollars from which it pays \$1000 to each employee who achieves high performance during the year. The probability of each employee achieving this goal is 0.10 and is independent of the probabilities of the other employees doing so. If there are $n = 10$ employees, how much should M equal so that the fund has a probability of at least 99% of covering those payments?

2.4-17. Your stockbroker is free to take your calls about 60% of the time; otherwise, he is talking to another client or is out of the office. You call him at five random times during a given month. (Assume independence.)

(a) What is the probability that he will take every one of the five calls?

(b) What is the probability that he will accept exactly three of your five calls?

(c) What is the probability that he will accept at least one of the calls?

2.4-18. In group testing for a certain disease, a blood sample was taken from each of n individuals and part of each sample was placed in a common pool. The latter was then tested. If the result was negative, there was no more testing and all n individuals were declared negative with one test. If, however, the combined result was found positive, all individuals were tested, requiring $n+1$ tests. If $p = 0.05$ is the probability of a person's having the disease and $n = 5$, compute the expected number of tests needed, assuming independence.

2.4-19. Define the pmf and give the values of μ, σ^2, and σ when the moment-generating function of X is defined by

(a) $M(t) = 1/3 + (2/3)e^t$.

(b) $M(t) = (0.25 + 0.75e^t)^{12}$.

2.4-20. **(i)** Give the name of the distribution of X (if it has a name), **(ii)** find the values of μ and σ^2, and **(iii)** calculate $P(1 \le X \le 2)$ when the moment-generating function of X is given by

(a) $M(t) = (0.3 + 0.7e^t)^5$.

(b) $M(t) = \dfrac{0.3e^t}{1 - 0.7e^t}, \qquad t < -\ln(0.7)$.

(c) $M(t) = 0.45 + 0.55e^t$.

(d) $M(t) = 0.3e^t + 0.4e^{2t} + 0.2e^{3t} + 0.1e^{4t}$.

(e) $M(t) = \sum_{x=1}^{10} (0.1)e^{tx}$.

2.5 THE NEGATIVE BINOMIAL DISTRIBUTION

We turn now to the situation in which we observe a sequence of independent Bernoulli trials until exactly r successes occur, where r is a fixed positive integer. Let the random variable X denote the number of trials needed to observe the rth success. That is, X is the trial number on which the rth success is observed. By the multiplication rule of probabilities, the pmf of X—say, $g(x)$—equals the product of the probability

$$\binom{x-1}{r-1}p^{r-1}(1-p)^{x-r} = \binom{x-1}{r-1}p^{r-1}q^{x-r}$$

of obtaining exactly $r-1$ successes in the first $x-1$ trials and the probability p of a success on the rth trial. Thus, the pmf of X is

$$g(x) = \binom{x-1}{r-1}p^{r}(1-p)^{x-r} = \binom{x-1}{r-1}p^{r}q^{x-r}, \quad x = r, r+1, \ldots.$$

We say that X has a **negative binomial distribution**.

REMARK The reason for calling this distribution the negative binomial distribution is as follows: Consider $h(w) = (1-w)^{-r}$, the binomial $(1-w)$ with the negative exponent $-r$. Using Maclaurin's series expansion, we have

$$(1-w)^{-r} = \sum_{k=0}^{\infty}\frac{h^{(k)}(0)}{k!}w^k = \sum_{k=0}^{\infty}\binom{r+k-1}{r-1}w^k, \qquad -1 < w < 1.$$

If we let $x = k + r$ in the summation, then $k = x - r$ and

$$(1-w)^{-r} = \sum_{x=r}^{\infty}\binom{r+x-r-1}{r-1}w^{x-r} = \sum_{x=r}^{\infty}\binom{x-1}{r-1}w^{x-r},$$

the summand of which is, except for the factor p^r, the negative binomial probability when $w = q$. In particular, the sum of the probabilities for the negative binomial distribution is 1 because

$$\sum_{x=r}^{\infty} g(x) = \sum_{x=r}^{\infty}\binom{x-1}{r-1}p^r q^{x-r} = p^r(1-q)^{-r} = 1.$$

If $r = 1$ in the negative binomial distribution, we note that X has a **geometric distribution**, since the pmf consists of terms of a geometric series, namely,

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

Recall that for a geometric series, the sum is given by

$$\sum_{k=0}^{\infty} ar^k = \sum_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r}$$

when $|r| < 1$. Thus, for the geometric distribution,

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

so that $g(x)$ does satisfy the properties of a pmf.

From the sum of a geometric series, we also note that when k is an integer,

$$P(X > k) = \sum_{x=k+1}^{\infty} (1-p)^{x-1} p = \frac{(1-p)^k p}{1-(1-p)} = (1-p)^k = q^k.$$

Thus, the value of the cdf at a positive integer k is

$$P(X \le k) = \sum_{x=1}^{k} (1-p)^{x-1} p = 1 - P(X > k) = 1 - (1-p)^k = 1 - q^k.$$

Example 2.5-1 Some biology students were checking eye color in a large number of fruit flies. For the individual fly, suppose that the probability of white eyes is 1/4 and the probability of red eyes is 3/4, and that we may treat these observations as independent Bernoulli trials. The probability that at least four flies have to be checked for eye color to observe a white-eyed fly is given by

$$P(X \ge 4) = P(X > 3) = q^3 = \left(\frac{3}{4}\right)^3 = \frac{27}{64} = 0.4219.$$

The probability that at most four flies have to be checked for eye color to observe a white-eyed fly is given by

$$P(X \le 4) = 1 - q^4 = 1 - \left(\frac{3}{4}\right)^4 = \frac{175}{256} = 0.6836.$$

The probability that the first fly with white eyes is the fourth fly considered is

$$P(X = 4) = q^{4-1} p = \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right) = \frac{27}{256} = 0.1055.$$

It is also true that

$$\begin{aligned} P(X = 4) &= P(X \le 4) - P(X \le 3) \\ &= [1 - (3/4)^4] - [1 - (3/4)^3] \\ &= \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right). \end{aligned}$$

■

We now show that the mean and the variance of a negative binomial random variable X are, respectively,

$$\mu = E(X) = \frac{r}{p} \quad \text{and} \quad \sigma^2 = \frac{rq}{p^2} = \frac{r(1-p)}{p^2}.$$

In particular, if $r = 1$, so that X has a geometric distribution, then

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

The mean $\mu = 1/p$ agrees with our intuition. Let's check: If $p = 1/6$, then we would expect, on the average, $1/(1/6) = 6$ trials before the first success.

To find these moments, we determine the mgf of the negative binomial distribution. It is

$$
\begin{aligned}
M(t) &= \sum_{x=r}^{\infty} e^{tx}\binom{x-1}{r-1}p^r(1-p)^{x-r} \\
&= (pe^t)^r \sum_{x=r}^{\infty}\binom{x-1}{r-1}\left[(1-p)e^t\right]^{x-r} \\
&= \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \qquad \text{where } (1-p)e^t < 1
\end{aligned}
$$

(or, equivalently, when $t < -\ln(1-p)$). Thus,

$$
\begin{aligned}
M'(t) &= (pe^t)^r(-r)[1-(1-p)e^t]^{-r-1}[-(1-p)e^t] \\
&\quad + r(pe^t)^{r-1}(pe^t)[1-(1-p)e^t]^{-r} \\
&= r(pe^t)^r[1-(1-p)e^t]^{-r-1}
\end{aligned}
$$

and

$$
\begin{aligned}
M''(t) &= r(pe^t)^r(-r-1)[1-(1-p)e^t]^{-r-2}[-(1-p)e^t] \\
&\quad + r^2(pe^t)^{r-1}(pe^t)[1-(1-p)e^t]^{-r-1}.
\end{aligned}
$$

Accordingly,

$$M'(0) = rp^rp^{-r-1} = rp^{-1}$$

and

$$
\begin{aligned}
M''(0) &= r(r+1)p^rp^{-r-2}(1-p) + r^2p^rp^{-r-1} \\
&= rp^{-2}[(1-p)(r+1)+rp] \;=\; rp^{-2}(r+1-p).
\end{aligned}
$$

Hence, we have

$$\mu = \frac{r}{p} \qquad \text{and} \qquad \sigma^2 = \frac{r(r+1-p)}{p^2} - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

Even these calculations are a little messy, so a somewhat easier way is given in Exercises 2.5-5 and 2.5-6.

Example 2.5-2

Suppose that during practice a basketball player can make a free throw 80% of the time. Furthermore, assume that a sequence of free-throw shooting can be thought of as independent Bernoulli trials. Let X equal the minimum number of free throws that this player must attempt to make a total of 10 shots. The pmf of X is

$$g(x) = \binom{x-1}{10-1}(0.80)^{10}(0.20)^{x-10}, \qquad x = 10, 11, 12, \ldots.$$

The mean, variance, and standard deviation of X are, respectively,

$$\mu = 10\left(\frac{1}{0.80}\right) = 12.5, \qquad \sigma^2 = \frac{10(0.20)}{0.80^2} = 3.125, \qquad \text{and} \qquad \sigma = 1.768.$$

And we have, for example,

$$P(X = 12) = g(12) = \binom{11}{9}(0.80)^{10}(0.20)^2 = 0.2362.$$

Example 2.5-3 To consider the effect of p and r on the negative binomial distribution, Figure 2.5-1 gives the probability histograms for four combinations of p and r. Note that since $r = 1$ in the first of these, it represents a geometric pmf.

When the moment-generating function exists, derivatives of all orders exist at $t = 0$. Thus, it is possible to represent $M(t)$ as a Maclaurin series, namely,

$$M(t) = M(0) + M'(0)\left(\frac{t}{1!}\right) + M''(0)\left(\frac{t^2}{2!}\right) + M'''(0)\left(\frac{t^3}{3!}\right) + \cdots.$$

If the Maclaurin series expansion of $M(t)$ exists and the moments are given, we can sometimes sum the Maclaurin series to obtain the closed form of $M(t)$. This approach is illustrated in the next example.

Example 2.5-4 Let the moments of X be defined by

$$E(X^r) = 0.8, \qquad r = 1, 2, 3, \ldots.$$

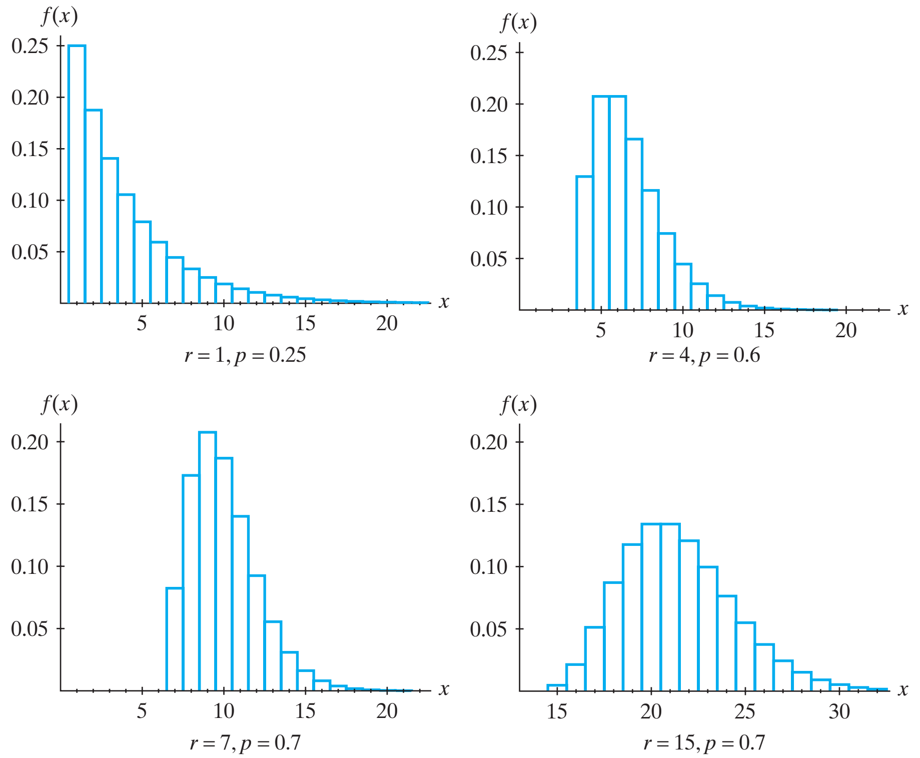

Figure 2.5-1 Negative binomial probability histograms

The moment-generating function of X is then

$$M(t) = M(0) + \sum_{r=1}^{\infty} 0.8\left(\frac{t^r}{r!}\right) = 1 + 0.8\sum_{r=1}^{\infty}\frac{t^r}{r!}$$

$$= 0.2 + 0.8\sum_{r=0}^{\infty}\frac{t^r}{r!} = 0.2e^{0t} + 0.8e^{1t}.$$

Thus,

$$P(X = 0) = 0.2 \qquad \text{and} \qquad P(X = 1) = 0.8.$$

This is an illustration of a Bernoulli distribution. ■

The next example gives an application of the geometric distribution.

Example 2.5-5 A fair six-sided die is rolled until each face is observed at least once. On the average, how many rolls of the die are needed? It always takes one roll to observe the first outcome. To observe a different face from the first roll is like observing a geometric random variable with $p = 5/6$ and $q = 1/6$. So on the average it takes $1/(5/6) = 6/5$ rolls. After two different faces have been observed, the probability of observing a new face is 4/6, so it will take, on the average, $1/(4/6) = 6/4$ rolls. Continuing in this manner, the answer is

$$1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = \frac{147}{10} = 14.7$$

rolls, on the average. ■

Exercises

2.5-1. An excellent free-throw shooter attempts several free throws until she misses.

(a) If $p = 0.9$ is her probability of making a free throw, what is the probability of having the first miss on the 13th attempt or later?

(b) If she continues shooting until she misses three, what is the probability that the third miss occurs on the 30th attempt?

2.5-2. Show that 63/512 is the probability that the fifth head is observed on the tenth independent flip of a fair coin.

2.5-3. Suppose that a basketball player different from the ones in Example 2.5-2 and in Exercise 2.5-1 can make a free throw 60% of the time. Let X equal the minimum number of free throws that this player must attempt to make a total of 10 shots.

(a) Give the mean, variance, and standard deviation of X.

(b) Find $P(X = 16)$.

2.5-4. Suppose an airport metal detector catches a person with metal 99% of the time. That is, it misses detecting a person with metal 1% of the time. Assume independence of people carrying metal. What is the probability that the first metal-carrying person missed (not detected) is among the first 50 metal-carrying persons scanned?

2.5-5. Let the moment-generating function $M(t)$ of X exist for $-h < t < h$. Consider the function $R(t) = \ln M(t)$. The first two derivatives of $R(t)$ are, respectively,

$$R'(t) = \frac{M'(t)}{M(t)} \qquad \text{and} \qquad R''(t) = \frac{M(t)M''(t) - [M'(t)]^2}{[M(t)]^2}.$$

Setting $t = 0$, show that

(a) $\mu = R'(0)$.

(b) $\sigma^2 = R''(0)$.

2.5-6. Use the result of Exercise 2.5-5 to find the mean and variance of the

(a) Bernoulli distribution.

(b) Binomial distribution.

(c) Geometric distribution.

(d) Negative binomial distribution.

2.5-7. If $E(X^r) = 5^r, r = 1, 2, 3, \ldots$, find the moment-generating function $M(t)$ of X and the pmf of X.

2.5-8. The probability that a company's work force has no accidents in a given month is 0.7. The numbers of accidents from month to month are independent. What is the probability that the third month in a year is the first month that at least one accident occurs?

2.5-9. One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased?

2.5-10. In 2012, Red Rose tea randomly began placing 1 of 12 English porcelain miniature figurines in a 100-bag box of the tea, selecting from 12 nautical figurines.

(a) On the average, how many boxes of tea must be purchased by a customer to obtain a complete collection consisting of the 12 nautical figurines?

(b) If the customer uses one tea bag per day, how long can a customer expect to take, on the average, to obtain a complete collection?

2.6 THE POISSON DISTRIBUTION

Some experiments result in counting the number of times particular events occur at given times or with given physical objects. For example, we could count the number of cell phone calls passing through a relay tower between 9 and 10 A.M., the number of flaws in 100 feet of wire, the number of customers that arrive at a ticket window between 12 noon and 2 P.M., or the number of defects in a 100-foot roll of aluminum screen that is 2 feet wide. Counting such events can be looked upon as observations of a random variable associated with an approximate Poisson process, provided that the conditions in the following definition are satisfied.

Definition 2.6-1

Let the number of occurrences of some event in a given continuous interval be counted. Then we have an **approximate Poisson process** with parameter $\lambda > 0$ if the following conditions are satisfied:

(a) The numbers of occurrences in nonoverlapping subintervals are independent.

(b) The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.

(c) The probability of two or more occurrences in a sufficiently short subinterval is essentially zero.

REMARK We use *approximate* to modify the Poisson process since we use *approximately* in (b) and *essentially* in (c) to avoid the "little o" notation. Occasionally, we simply say "Poisson process" and drop *approximate*.

Suppose that an experiment satisfies the preceding three conditions of an approximate Poisson process. Let X denote the number of occurrences in an interval of length 1 (where "length 1" represents one unit of the quantity under consideration). We would like to find an approximation for $P(X = x)$, where x is a nonnegative integer. To achieve this, we partition the unit interval into n subintervals of equal length $1/n$. If n is sufficiently large (i.e., much larger than x), we shall approximate the probability that there are x occurrences in this unit interval by finding the probability that exactly x of these n subintervals each has one occurrence. The probability of one occurrence in any one subinterval of length $1/n$ is approximately $\lambda(1/n)$, by condition (b). The probability of two or more occurrences in any one subinterval is essentially zero, by condition (c). So, for each subinterval, there is exactly one occurrence with a probability of approximately $\lambda(1/n)$. Consider the occurrence or nonoccurrence in each subinterval as a Bernoulli trial. By condition (a), we have a sequence of n Bernoulli trials with probability p approximately equal to $\lambda(1/n)$. Thus, an approximation for $P(X = x)$ is given by the binomial probability

$$\frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

If n increases without bound, then

$$\begin{aligned}\lim_{n\to\infty} &\frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x}\,\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}.\end{aligned}$$

Now, for fixed x, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{n(n-1)\cdots(n-x+1)}{n^x} &= \lim_{n\to\infty}\left[(1)\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right] = 1, \\ \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n &= e^{-\lambda}, \\ \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} &= 1.\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!} = P(X = x).$$

The distribution of probability associated with this process has a special name. We say that the random variable X has a **Poisson distribution** if its pmf is of the form

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

where $\lambda > 0$.

It is easy to see that $f(x)$ has the properties of a pmf because, clearly, $f(x) \geq 0$ and, from the Maclaurin series expansion of e^{λ}, we have

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

To discover the exact role of the parameter $\lambda > 0$, let us find some of the characteristics of the Poisson distribution. The mgf of X is

$$M(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}.$$

From the series representation of the exponential function, we have

$$M(t) = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

for all real values of t. Now,

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

and

$$M''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}.$$

The values of the mean and variance of X are, respectively,

$$\mu = M'(0) = \lambda$$

and

$$\sigma^2 = M''(0) - [M'(0)]^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

That is, for the Poisson distribution, $\mu = \sigma^2 = \lambda$.

REMARK It is also possible to find the mean and the variance for the Poisson distribution directly, without using the mgf. The mean for the Poisson distribution is given by

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}$$

because $(0)f(0) = 0$ and $x/x! = 1/(x-1)!$ when $x > 0$. If we let $k = x - 1$, then

$$\begin{aligned} E(X) &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} = \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= \lambda\, e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

To find the variance, we first determine the second factorial moment $E[X(X-1)]$. We have

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!}$$

because $(0)(0-1)f(0) = 0$, $(1)(1-1)f(1) = 0$, and $x(x-1)/x! = 1/(x-2)!$ when $x > 1$. If we let $k = x - 2$, then

$$\begin{aligned} E[X(X-1)] &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+2}}{k!} = \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2. \end{aligned}$$

Thus,

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

We again see that, for the Poisson distribution, $\mu = \sigma^2 = \lambda$. ■

Table III in Appendix B gives values of the cdf of a Poisson random variable for selected values of λ. This table is illustrated in the next example.

Example 2.6-1 Let X have a Poisson distribution with a mean of $\lambda = 5$. Then, using Table III in Appendix B, we obtain

$$P(X \le 6) = \sum_{x=0}^{6} \frac{5^x e^{-5}}{x!} = 0.762,$$

$$P(X > 5) = 1 - P(X \le 5) = 1 - 0.616 = 0.384,$$

and

$$P(X = 6) = P(X \le 6) - P(X \le 5) = 0.762 - 0.616 = 0.146.$$

Example 2.6-2 To see the effect of λ on the pmf $f(x)$ of X, Figure 2.6-1 shows the probability histograms of $f(x)$ for four different values of λ.

If events in an approximate Poisson process occur at a mean rate of λ per unit, then the expected number of occurrences in an interval of length t is λt. For example, let X equal the number of alpha particles emitted by barium-133 in one second and counted by a Geiger counter. If the mean number of emitted particles is 60 per second, then the expected number of emitted particles in 1/10 of a second is $60(1/10) = 6$. Moreover, the number of emitted particles, say X, in a time interval of length t has the Poisson pmf

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

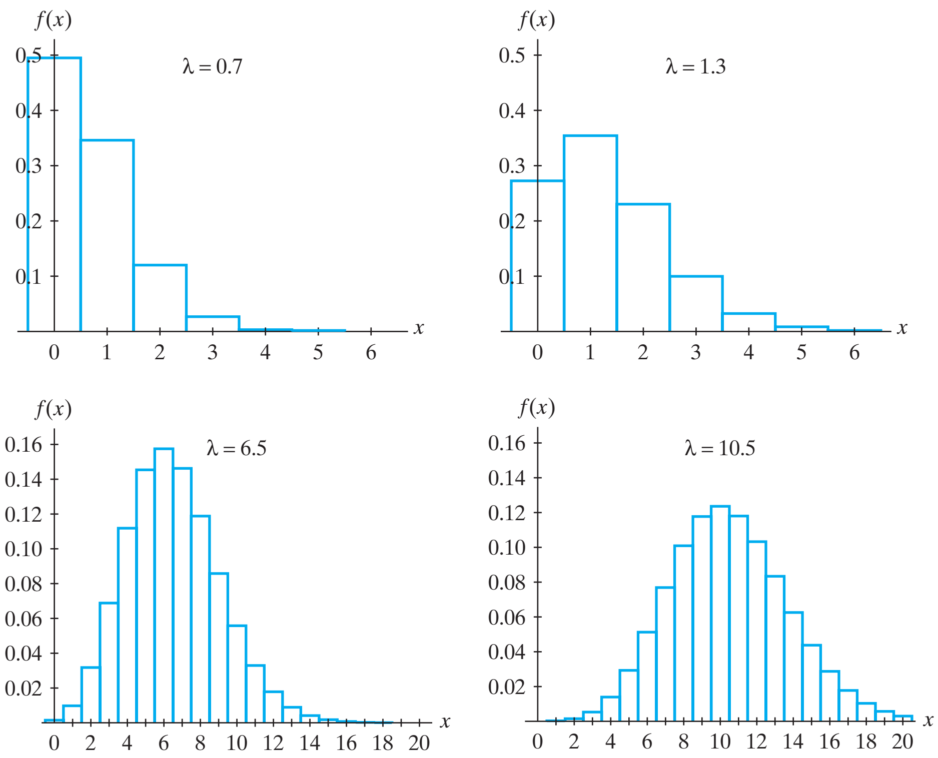

Figure 2.6-1 Poisson probability histograms

This equation follows if we treat the interval of length t as if it were the "unit interval" with mean λt instead of λ.

Example 2.6-3

A USB flash drive is sometimes used to back up computer files. However, in the past, a less reliable backup system that was used was a computer tape, and flaws occurred on these tapes. In a particular situation, flaws (bad records) on a used computer tape occurred on the average of one flaw per 1200 feet. If one assumes a Poisson distribution, what is the distribution of X, the number of flaws in a 4800-foot roll? The expected number of flaws in 4800 = 4(1200) feet is 4; that is, $E(X) = 4$. Thus, the pmf of X is

$$f(x) = \frac{4^x e^{-4}}{x!}, \qquad x = 0, 1, 2, \ldots,$$

and, in particular,

$$P(X = 0) = \frac{4^0 e^{-4}}{0!} = e^{-4} = 0.018,$$
$$P(X \leq 4) = 0.629,$$

by Table III in Appendix B. ■

Example 2.6-4

In a large city, telephone calls to 911 come on the average of two every 3 minutes. If one assumes an approximate Poisson process, what is the probability of five or more calls arriving in a 9-minute period? Let X denote the number of calls in a 9-minute period. We see that $E(X) = 6$; that is, on the average, six calls will arrive during a 9-minute period. Thus,

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \frac{6^x e^{-6}}{x!}$$
$$= 1 - 0.285 = 0.715,$$

by Table III in Appendix B. ■

Not only is the Poisson distribution important in its own right, but it can also be used to approximate probabilities for a binomial distribution. Earlier we saw that if X has a Poisson distribution with parameter λ, then with n large,

$$P(X = x) \approx \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x},$$

where $p = \lambda/n$, so that $\lambda = np$ in the above binomial probability. That is, if X has the binomial distribution $b(n, p)$ with large n and small p, then

$$\frac{(np)^x e^{-np}}{x!} \approx \binom{n}{x} p^x (1 - p)^{n-x}.$$

This approximation is reasonably good if n is large. But since λ was a fixed constant in that earlier argument, p should be small, because $np = \lambda$. In particular, the approximation is quite accurate if $n \geq 20$ and $p \leq 0.05$ or if $n \geq 100$ and $p \leq 0.10$, but it is not bad in other situations violating these bounds somewhat, such as $n = 50$ and $p = 0.12$.

Example 2.6-5

A manufacturer of Christmas tree light bulbs knows that 2% of its bulbs are defective. Assuming independence, the number of defective bulbs in a box of 100 bulbs has a binomial distribution with parameters $n = 100$ and $p = 0.02$. To approximate the probability that a box of 100 of these bulbs contains at most three defective bulbs, we use the Poisson distribution with $\lambda = 100(0.02) = 2$, which gives

$$\sum_{x=0}^{3} \frac{2^x e^{-2}}{x!} = 0.857,$$

from Table III in Appendix B. Using the binomial distribution, we obtain, after some tedious calculations,

$$\sum_{x=0}^{3} \binom{100}{x} (0.02)^x (0.98)^{100-x} = 0.859.$$

Hence, in this case, the Poisson approximation is extremely close to the true value, but much easier to find. ■

REMARK With the availability of statistical computer packages and statistical calculators, it is often very easy to find binomial probabilities. So do not use the Poisson approximation if you are able to find the probability exactly. ■

Example 2.6-6

In Figure 2.6-2, Poisson probability histograms have been superimposed on shaded binomial probability histograms so that we can see whether or not these are close to

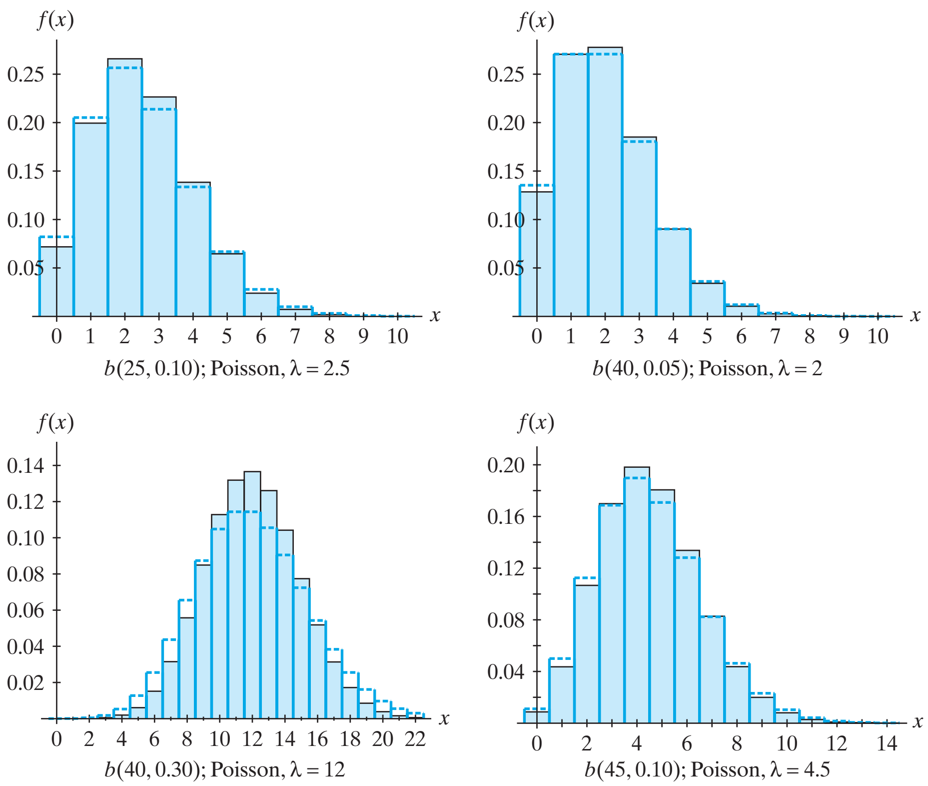

Figure 2.6-2 Binomial (shaded) and Poisson probability histograms

each other. If the distribution of X is $b(n, p)$, the approximating Poisson distribution has a mean of $\lambda = np$. Note that the approximation is not good when p is large (e.g., $p = 0.30$). ■

Exercises

2.6-1. Let X have a Poisson distribution with a mean of 4. Find

(a) $P(2 \le X \le 5)$.

(b) $P(X \ge 3)$.

(c) $P(X \le 3)$.

2.6-2. Let X have a Poisson distribution with a variance of 3. Find $P(X = 2)$.

2.6-3. Customers arrive at a travel agency at a mean rate of 11 per hour. Assuming that the number of arrivals per hour has a Poisson distribution, give the probability that more than 10 customers arrive in a given hour.

2.6-4. Find $P(X = 4)$ if X has a Poisson distribution such that $3P(X = 1) = P(X = 2)$.

2.6-5. Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume a Poisson distribution, find the probability of at most one flaw appearing in 225 square feet.

2.6-6. A certain type of aluminum screen that is 2 feet wide has, on the average, one flaw in a 100-foot roll. Find the probability that a 50-foot roll has no flaws.

2.6-7. With probability 0.001, a prize of \$499 is won in the Michigan Daily Lottery when a \$1 straight bet is placed. Let Y equal the number of \$499 prizes won by a gambler after placing n straight bets. Note that Y is $b(n, 0.001)$. After placing $n = 2000$ \$1 bets, the gambler is behind or even if $\{Y \le 4\}$. Use the Poisson distribution to approximate $P(Y \le 4)$ when $n = 2000$.

2.6-8. Suppose that the probability of suffering a side effect from a certain flu vaccine is 0.005. If 1000 persons are inoculated, find the approximate probability that

(a) At most 1 person suffers.

(b) 4, 5, or 6 persons suffer.

2.6-9. A store selling newspapers orders only $n = 4$ of a certain newspaper because the manager does not get many calls for that publication. If the number of requests per day follows a Poisson distribution with mean 3,

(a) What is the expected value of the number sold?

(b) What is the minimum number that the manager should order so that the chance of having more requests than available newspapers is less than 0.05?

2.6-10. The mean of a Poisson random variable X is $\mu = 9$. Compute

$$P(\mu - 2\sigma < X < \mu + 2\sigma).$$

2.6-11. An airline always overbooks if possible. A particular plane has 95 seats on a flight in which a ticket sells for \$300. The airline sells 100 such tickets for this flight.

(a) If the probability of an individual not showing up is 0.05, assuming independence, what is the probability that the airline can accommodate all the passengers who do show up?

(b) If the airline must return the \$300 price plus a penalty of \$400 to each passenger that cannot get on the flight, what is the expected payout (penalty plus ticket refund) that the airline will pay?

2.6-12. A baseball team loses \$100,000 for each consecutive day it rains. Say X, the number of consecutive days it rains at the beginning of the season, has a Poisson distribution with mean 0.2. What is the expected loss before the opening game?

2.6-13. Assume that a policyholder is four times more likely to file exactly two claims as to file exactly three claims. Assume also that the number X of claims of this policyholder is Poisson. Determine the expectation $E(X^2)$.

HISTORICAL COMMENTS The next major items advanced in probability theory were by the Bernoullis, a remarkable Swiss family of mathematicians of the late 1600s to the late 1700s. There were eight mathematicians among them, but we shall mention just three of them: Jacob, Nicolaus II, and Daniel. While writing *Ars Conjectandi* (*The Art of Conjecture*), Jacob died in 1705, and a nephew, Nicolaus II, edited the work for publication. However, it was Jacob who discovered the important law of large numbers, which is included in our Section 5.8.

Another nephew of Jacob, Daniel, noted in his St. Petersburg paper that "expected values are computed by multiplying each possible gain by the number of ways in which it can occur and then dividing the sum of these products by the total number of cases." His cousin, Nicolaus II, then proposed the so-called St. Petersburg paradox: Peter continues to toss a coin until a head first appears—say, on the xth trial—and he then pays Paul 2^{x-1} units (originally ducats, but for convenience we use dollars). With each additional throw, the number of dollars has doubled. How much should another person pay Paul to take his place in this game? Clearly,

$$E(2^{X-1}) = \sum_{x=1}^{\infty} (2^{x-1})\left(\frac{1}{2^x}\right) = \sum_{x=1}^{\infty} \frac{1}{2} = \infty.$$

However, if we consider this as a practical problem, would someone be willing to give Paul \$1000 to take his place even though there is this unlimited expected value? We doubt it and Daniel doubted it, and it made him think about the utility of money. For example, to most of us, \$3 million is not worth three times \$1 million. To convince you of that, suppose you had exactly \$1 million and a very rich man offers to bet you \$2 million against your \$1 million on the flip of a coin. You will have zero or \$3 million after the flip, so your expected value is

$$(\$0)\left(\frac{1}{2}\right) + (\$3{,}000{,}000)\left(\frac{1}{2}\right) = \$1{,}500{,}000,$$

much more than your \$1 million. Seemingly, then, this is a great bet and one that Bill Gates might take. However, remember you have \$1 million for certain and you could have zero with probability 1/2. None of us with limited resources should consider taking that bet, because the utility of that extra money to us is not worth the utility of the first \$1 million. Now, each of us has our own utility function. Two dollars is worth twice as much as one dollar for practically all of us. But is \$200,000 worth twice as much as \$100,000? It depends upon your situation; so while the utility function is a straight line for the first several dollars, it still increases but begins to bend downward someplace as the amount of money increases. This occurs at different spots for all of us. Bob Hogg, one of the authors of this text, would bet \$1000 against \$2000 on a flip of the coin anytime, but probably not \$100,000 against \$200,000, so Hogg's utility function has started to bend downward someplace between \$1000 and \$100,000. Daniel Bernoulli made this observation, and it is extremely useful in all kinds of businesses.

As an illustration, in insurance, most of us know that the premium we pay for all types of insurance is greater than what the company expects to pay us; that is how they make money. Seemingly, insurance is a bad bet, but it really isn't always. It is true that we should self-insure less expensive items—those whose value is on that straight part of the utility function. We have even heard the "rule" that you not insure anything worth less than two months' salary; this is a fairly good guide, but each of us has our own utility function and must make that decision. Hogg can afford losses in the \$5000 to \$10,000 range (not that he likes them, of course), but he does not want to pay losses of \$100,000 or more. So his utility function for negative values of the argument follows that straight line for relatively small negative amounts but again bends down for large negative amounts. If you insure expensive items, you will discover that the expected utility in absolute value will now exceed the premium. This is why most people insure their life, their home, and their car (particularly on the liability side). They should not, however, insure their golf clubs, eyeglasses, furs, or jewelry (unless the latter two items are extremely valuable).