Discrete Random Variables and Probability Distributions

2

Suppose a city's traffic engineering department monitors a certain intersection during a one-hour period in the middle of the day. Many characteristics might be of interest to the observers, including the number of vehicles that enter the intersection, the largest number of vehicles in the left turn lane during a signal cycle, the speed of the fastest vehicle going through the intersection, the average speed of all vehicles entering the intersection. The value of each one of the foregoing variable quantities is subject to uncertainty—we don't know a priori how many vehicles will enter, what the maximum speed will be, etc. So each of these is referred to as a *random variable*—a variable quantity whose value is determined by what happens in a chance experiment.

There are two fundamentally different types of random variables, discrete and continuous. In this chapter we examine the basic properties and introduce the most important examples of discrete random variables. Chapter 3 covers the same territory for continuous random variables.

2.1   Random Variables

In any experiment, numerous characteristics can be observed or measured, but in most cases an experimenter will focus on some specific aspect or aspects of a sample. For example, in a study of commuting patterns in a metropolitan area, each individual in a sample might be asked about commuting distance and the number of people commuting in the same vehicle, but not about IQ, income, family size, and other such characteristics. Alternatively, a researcher may test a sample of components and record only the number that have failed within 1000 hours, rather than record the individual failure times.

In general, each outcome of an experiment can be associated with a number by specifying a rule of association (e.g., the number among the sample of ten components that fail to last 1,000 h or the total weight of baggage for a sample of 25 airline passengers). Such a rule of association is called a **random variable**—a variable because different numerical values are possible and random because the observed value depends on which of the possible experimental outcomes results (Fig. 2.1).
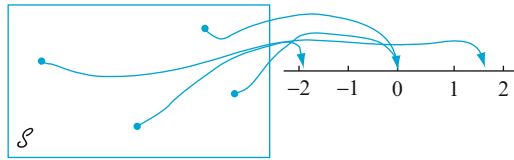
Fig. 2.1 A random variable

DEFINITION

For a given sample space $\mathscr{S}$ of some experiment, a **random variable** (**rv**) is any rule that associates a number with each outcome in $\mathscr{S}$. In mathematical language, a random variable is a function whose domain is the sample space and whose range is some subset of real numbers.

Random variables are customarily denoted by uppercase letters, such as X and Y, near the end of our alphabet. We will use lowercase letters to represent some particular value of the corresponding random variable. The notation $X(s) = x$ means that x is the value associated with the outcome s by the rv X.

Example 2.1 When a student attempts to connect to a university computer system, either there is a failure (F) or there is a success (S). With $\mathscr{S} = \{S, F\}$, define an rv X by $X(S) = 1$, $X(F) = 0$. The rv X indicates whether (1) or not (0) the student can connect.   ■

In Example 2.1, the rv X was specified by explicitly listing each element of $\mathscr{S}$ and the associated number. If $\mathscr{S}$ contains more than a few outcomes, such a listing is tedious, but it can frequently be avoided.

Example 2.2 Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer (such devices are used extensively by polling organizations), and define an rv Y by

$$Y = \begin{cases} 1 & \text{if the selected number is unlisted} \\ 0 & \text{if the selected number is listed in the directory} \end{cases}$$

For example, if 5282966 appears in the telephone directory, then $Y(5282966) = 0$, whereas $Y(7727350) = 1$ tells us that the number 7727350 is unlisted. A word description of this sort is more economical than a complete listing, so we will use such a description whenever possible.   ■

In Examples 2.1 and 2.2, the only possible values of the random variable were 0 and 1. Such a random variable arises frequently enough to be given a special name, after the individual who first studied it.

DEFINITION

Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable**.

We will often want to define and study several different random variables from the same sample space.

Example 2.3 Example 1.3 described an experiment in which the number of pumps in use at each of two gas stations was determined. Define rvs X, Y, and U by

$X = $ the total number of pumps in use at the two stations
$Y = $ the difference between the number of pumps in use at station 1 and the number in use at station 2
$U = $ the maximum of the numbers of pumps in use at the two stations

If this experiment is performed and $s = (2, 3)$ results, then $X((2, 3)) = 2 + 3 = 5$, so we say that the observed value of X is $x = 5$. Similarly, the observed value of Y would be $y = 2 - 3 = -1$, and the observed value of U would be $u = \max(2, 3) = 3$.   ∎

Each of the random variables of Examples 2.1–2.3 can assume only a finite number of possible values. This need not be the case.

Example 2.4 Consider an experiment in which 9-V batteries are examined until one with an acceptable voltage (S) is obtained. The sample space is $\mathcal{S} = \{S, FS, FFS, \dots\}$. Define an rv X by

$X = $ the number of batteries examined before the experiment terminates

Then $X(S) = 1, X(FS) = 2, X(FFS) = 3, \dots, X(FFFFFFS) = 7$, and so on. Any positive integer is a possible value of X, so the set of possible values is infinite.   ∎

Example 2.5 Suppose that in some random fashion, a location (latitude and longitude) in the continental USA is selected. Define an rv Y by

$Y = $ the height, in feet, above sea level at the selected location

For example, if the selected location were $(39°50'N, 98°35'W)$, then we might have $Y((39°50'N, 98°35'W)) = 1748.26$ ft. The largest possible value of Y is 14,494 (Mt. Whitney), and the smallest possible value is -282 (Death Valley). The set of all possible values of Y is the set of all numbers in the interval between -282 and 14,494; that is, the range of Y is

$$\{y : -282 \leq y \leq 14,494\} = [-282, 14,494]$$

and there are infinitely-many numbers in this interval.   ∎

2.1.1   Two Types of Random Variables

Determining the values of variables such as the number of visits to a website during a 24-h period or the number of patients in an emergency room at a particular time requires only counting. On the other hand, determining values of variables such as fuel efficiency of a vehicle (mpg) or reaction time to a stimulus necessitates making a measurement of some sort. The following definition formalizes the distinction between these two different kinds of variables.

> **DEFINITION**
>
> A **discrete** random variable is an rv whose possible values constitute either a finite set or a countably infinite set (e.g., the set of all integers, or the set of all positive integers).
>
> A random variable is **continuous** if *both* of the following apply:
> 1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).
> 2. No possible value of the variable has positive probability, that is, $P(X = c) = 0$ for any possible value c.

Although any interval on the number line contains infinitely-many numbers, it can be shown that there is no way to create a listing of all these values—there are just too many of them. The second condition describing a continuous random variable is perhaps counterintuitive, since it would seem to imply a total probability of zero for all possible values. But we shall see in Chap. 3 that *intervals* of values have positive probability; the probability of an interval will decrease to zero as the width of the interval shrinks to zero. In practice, discrete variables virtually always involve counting the number of something, whereas continuous variables entail making measurements of some sort.

Example 2.6 All random variables in Examples 2.1–2.4 are discrete. As another example, suppose we select married couples at random and do a blood test on each person until we find a husband and wife who both have the same Rh factor. With $X =$ the number of blood tests to be performed, possible values of X are $\{2, 4, 6, 8, \ldots\}$. Since the possible values have been listed in sequence, X is a discrete rv. ∎

To study basic properties of discrete rvs, only the tools of discrete mathematics—summation and differences—are required. The study of continuous variables in Chap. 3 will require the continuous mathematics of the calculus—integrals and derivatives.

2.1.2 Exercises: Section 2.1 (1–10)

1. A concrete beam may fail either by shear (S) or flexure (F). Suppose that three failed beams are randomly selected and the type of failure is determined for each one. Let $X =$ the number of beams among the three selected that failed by shear. List each outcome in the sample space along with the associated value of X.
2. Give three examples of Bernoulli rvs (other than those in the text).
3. Using the experiment in Example 2.3, define two more random variables and list the possible values of each.
4. Let $X =$ the number of nonzero digits in a randomly selected zip code. What are the possible values of X? Give three possible outcomes and their associated X values.
5. If the sample space S is an infinite set, does this necessarily imply that any rv X defined from S will have an infinite set of possible values? If yes, say why. If no, give an example.
6. Starting at a fixed time, each car entering an intersection is observed to see whether it turns left (L), right (R), or goes straight ahead (A). The experiment terminates as soon as a car is observed to turn left. Let $X =$ the number of cars observed. What are possible X values? List five outcomes and their associated X values.

7. For each random variable defined here, describe the set of possible values for the variable, and state whether the variable is discrete.
   (a) X = the number of unbroken eggs in a randomly chosen standard egg carton
   (b) Y = the number of students on a class list for a particular course who are absent on the first day of classes
   (c) U = the number of times a duffer has to swing at a golf ball before hitting it
   (d) X = the length of a randomly selected rattlesnake
   (e) Z = the amount of royalties earned from the sale of a first edition of 10,000 textbooks
   (f) Y = the acidity level (pH) of a randomly chosen soil sample
   (g) X = the tension (psi) at which a randomly selected tennis racket has been strung
   (h) X = the total number of coin tosses required for three individuals to obtain a match (HHH or TTT)
8. Each time a component is tested, the trial is a success (S) or failure (F). Suppose the component is tested repeatedly until a success occurs on three *consecutive* trials. Let Y denote the number of trials necessary to achieve this. List all outcomes corresponding to the five smallest possible values of Y, and state which Y value is associated with each one.
9. An individual named Claudius is located at the point 0 in the accompanying diagram.
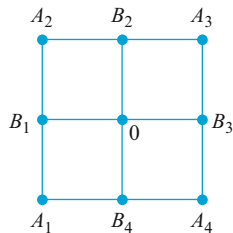
Using an appropriate randomization device (such as a tetrahedral die, one having four sides), Claudius first moves to one of the four locations B_1, B_2, B_3, B_4. Once at one of these locations, he uses another randomization device to decide whether he next returns to 0 or next visits one of the other two adjacent points. This process then continues; after each move, another move to one of the (new) adjacent points is determined by tossing an appropriate die or coin.
   (a) Let X = the number of moves that Claudius makes before first returning to 0. What are possible values of X? Is X discrete or continuous?
   (b) If moves are allowed also along the diagonal paths connecting 0 to A_1, A_2, A_3, and A_4, respectively, answer the questions in part (a).
10. The number of pumps in use at both a six-pump station and a four-pump station will be determined. Give the possible values for each of the following random variables:
   (a) T = the total number of pumps in use
   (b) X = the difference between the numbers in use at stations 1 and 2
   (c) U = the maximum number of pumps in use at either station
   (d) Z = the number of stations having exactly two pumps in use

2.2 Probability Distributions for Discrete Random Variables

When probabilities are assigned to various outcomes in $\mathscr{S}$, these in turn determine probabilities associated with the values of any particular rv X. The *probability distribution of X* says how the total probability of 1 is distributed among (allocated to) the various possible X values.

Example 2.7 Six batches of components are ready to be shipped by a supplier. The number of defective components in each batch is as follows:

Batch	#1	#2	#3	#4	#5	#6
Number of defectives	0	2	0	1	2	0

One of these batches is to be randomly selected for shipment to a customer. Let X be the number of defectives in the selected batch. The three possible X values are 0, 1, and 2. Of the six equally likely simple events, three result in $X = 0$, one in $X = 1$, and the other two in $X = 2$. Let $p(0)$ denote the probability that $X = 0$ and $p(1)$ and $p(2)$ represent the probabilities of the other two possible values of X. Then

$$p(0) = P(X = 0) = P(\text{batch 1 or 3 or 6 is sent}) = \frac{3}{6} = .500$$

$$p(1) = P(X = 1) = P(\text{batch 4 is sent}) = \frac{1}{6} = .167$$

$$p(2) = P(X = 2) = P(\text{batch 2 or 5 is sent}) = \frac{2}{6} = .333$$

That is, a probability of .500 is distributed to the X value 0, a probability of .167 is placed on the X value 1, and the remaining probability, .333, is associated with the X value 2. The values of X along with their probabilities collectively specify the probability distribution or *probability mass function of X*. If this experiment were repeated over and over again, in the long run $X = 0$ would occur one-half of the time, $X = 1$ one-sixth of the time, and $X = 2$ one-third of the time. ∎

DEFINITION

The **probability distribution** or **probability mass function** (pmf) of a discrete rv is defined for every number x by

$$p(x) = P(X = x) = P(\text{all } s \in \mathcal{S}: X(s) = x).^{[1]}$$

In words, for every possible value x of the random variable, the pmf specifies the probability of observing that value when the experiment is performed. The conditions $p(x) \geq 0$ and $\Sigma p(x) = 1$, where the summation is over all possible x, are required of any pmf.

Example 2.8 Consider randomly selecting a student at a large public university, and define a Bernoulli rv by $X = 1$ if the selected student does not qualify for in-state tuition (a success from the university administration's point of view) and $X = 0$ if the student does qualify. If 20% of all students do not qualify, the pmf for X is

$p(0) = P(X = 0) = P(\text{the selected student does qualify}) = .8$
$p(1) = P(X = 1) = P(\text{the selected student does not qualify}) = .2$
$p(x) = P(X = x) = 0 \text{ for } x \neq 0 \text{ or } 1.$

---

[1] $P(X = x)$ is read "the probability that the rv X assumes the value x." For example, $P(X = 2)$ denotes the probability that the resulting X value is 2.

$$p(x) = \begin{cases} .8 & \text{if } x = 0 \\ .2 & \text{if } x = 1 \\ 0 & \text{if } x \neq 0 \text{ or } 1 \end{cases}$$

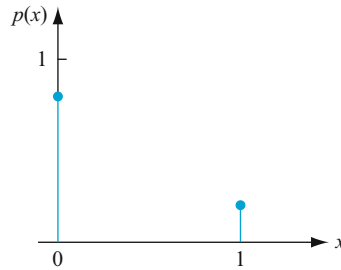Figure 2.2 is a picture of this pmf, called a *line graph*.

Fig. 2.2 The line graph for the pmf in Example 2.8 ∎

Example 2.9 Consider a group of five potential blood donors—A, B, C, D, and E—of whom only A and B have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified. Let the rv $Y =$ the number of typings necessary to identify an O+ individual. Then the pmf of Y is

$$p(1) = P(Y = 1) = P(\text{A or B typed first}) = \frac{2}{5} = .4$$
$$p(2) = P(Y = 2) = P(\text{C, D, or E first, and then A or B})$$
$$\qquad = P(\text{C, D, or E first}) \cdot P(\text{A or B next}\,|\, \text{C, D, or E first}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$
$$p(3) = P(Y = 3) = P(\text{C, D, or E first and second, and then A or B}) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = .2$$
$$p(4) = P(Y = 4) = P(\text{C, D, and E all done first}) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = .1$$
$$p(y) = 0 \text{ for } y \neq 1, 2, 3, 4.$$

The pmf can be presented compactly in tabular form:

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

where any y value not listed receives zero probability. Figure 2.3 shows the line graph for this pmf.
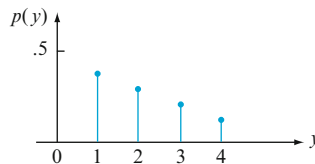

Fig. 2.3 The line graph for the pmf in Example 2.9 ∎

The name "probability mass function" is suggested by a model used in physics for a system of "point masses." In this model, masses are distributed at various locations x along a one-dimensional axis. Our pmf describes how the total probability mass of 1 is distributed at various points along the axis of possible values of the random variable (where and how much mass at each x).

Another useful pictorial representation of a pmf is called a **probability histogram**. Above each y with $p(y) > 0$, construct a rectangle centered at y. The height of each rectangle is proportional to $p(y)$, and the base is the same for all rectangles. When possible values are equally spaced, the base is frequently chosen as the distance between successive y values (though it could be smaller). Figure 2.4 shows two probability histograms.
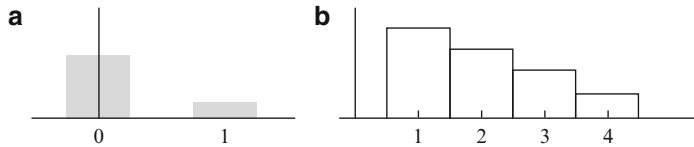

Fig. 2.4 Probability histograms: (**a**) Example 2.8; (**b**) Example 2.9

2.2.1 A Parameter of a Probability Distribution

In Example 2.8, we had $p(0) = .8$ and $p(1) = .2$. At another university, it may be the case that $p(0) = .9$ and $p(1) = .1$. More generally, the pmf of any Bernoulli rv can be expressed in the form $p(1) = \alpha$ and $p(0) = 1 - \alpha$, where $0 < \alpha < 1$. Because the pmf depends on the particular value of α, we often write $p(x; \alpha)$ rather than just $p(x)$:

$$p(x; \alpha) = \begin{cases} 1 - \alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.1)$$

Then each choice of α in Expression (2.1) yields a different pmf.

DEFINITION

Suppose $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

The quantity α in Expression (2.1) is a parameter. Each different number α between 0 and 1 determines a different member of a family of distributions; two such members are

$$p(x; .6) = \begin{cases} .4 & \text{if } x = 0 \\ .6 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad p(x; .5) = \begin{cases} .5 & \text{if } x = 0 \\ .5 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Every probability distribution for a Bernoulli rv has the form of Expression (2.1), so it is called the *family of Bernoulli distributions*.

Example 2.10 Starting at a fixed time, we observe the gender of each newborn child at a certain hospital until a boy (B) is born. Let $p = P(B)$, assume that successive births are independent, and define the rv X by $X =$ number of births observed. Then

$$p(1) = P(X = 1) = P(B) = p$$
$$p(2) = P(X = 2) = P(GB) = P(G) \cdot P(B) = (1 - p)p$$

and

$$p(3) = P(X = 3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1 - p)^2 p$$

Continuing in this way, a general formula emerges:

$$p(x) = \begin{cases} (1 - p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

The quantity p in Expression (2.2) represents a number between 0 and 1 and is a parameter of the probability distribution. In the gender example, $p = .51$ might be appropriate, but if we were looking for the first child with Rh-positive blood, then we might have $p = .85$. The random variable X has what is known as a *geometric distribution*, which we will discuss in Sect. 2.6.   ∎

2.2.2   The Cumulative Distribution Function

For some fixed value x, we often wish to compute the probability that the observed value of X will be *at most* x. For example, let X be the number of beds occupied in a hospital's emergency room at a certain time of day, and suppose the pmf of X is given by

x	0	1	2	3	4
$p(x)$	.20	.25	.30	.15	.10

Then the probability that at most two beds are occupied is $P(X \leq 2) = p(0) + p(1) + p(2) = .75$. Furthermore, since $X \leq 2.7$ iff $X \leq 2$, we also have $P(X \leq 2.7) = .75$, and similarly $P(X \leq 2.999) = .75$. Since 0 is the smallest possible X value, $P(X \leq -1.5) = 0$, $P(X \leq -10) = 0$, and in fact for any negative number x, $P(X \leq x) = 0$. And because 4 is the largest possible value of X, $P(X \leq 4) = 1$, $P(X \leq 9.8) = 1$, and so on.

Very importantly, $P(X < 2) = p(0) + p(1) = .45 < .75 = P(X \leq 2)$, because the latter probability includes the probability mass at the x value 2 whereas the former probability does not. More generally, $P(X < x) < P(X \leq x)$ whenever x is a possible value of X. Furthermore, $P(X \leq x)$ is a well-defined and computable probability for *any* number x.

DEFINITION
The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y : y \leq x} p(y) \tag{2.3}$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x.

Example 2.11 A store carries flash drives with 1, 2, 4, 8, or 16 GB of memory. The accompanying table gives the distribution of $X =$ the amount of memory in a purchased drive:

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

Let's first determine $F(x)$ for each of the five possible values of X:

$$F(1) = P(X \leq 1) = P(X = 1) = p(1) = .05$$
$$F(2) = P(X \leq 2) = P(X = 1 \text{ or } 2) = p(1) + p(2) = .15$$
$$F(4) = P(X \leq 4) = P(X = 1 \text{ or } 2 \text{ or } 4) = p(1) + p(2) + p(4) = .50$$
$$F(8) = P(X \leq 8) = p(1) + p(2) + p(4) + p(8) = .90$$
$$F(16) = P(X \leq 16) = 1$$

Now for any other number x, $F(x)$ will equal the value of F at the closest possible value of X to the left of x. For example,

$$F(2.7) = P(X \leq 2.7) = P(X \leq 2) = F(2) = .15$$
$$F(7.999) = P(X \leq 7.999) = P(X \leq 4) = F(4) = .50$$

If x is less than 1, $F(x) = 0$ [e.g., $F(.58) = 0$], and if x is at least 16, $F(x) = 1$ [e.g., $F(25) = 1$]. The cdf is thus

$$F(x) = \begin{cases} 0 & x < 1 \\ .05 & 1 \leq x < 2 \\ .15 & 2 \leq x < 4 \\ .50 & 4 \leq x < 8 \\ .90 & 8 \leq x < 16 \\ 1 & 16 \leq x \end{cases}$$

A graph of this cdf is shown in Fig. 2.5.
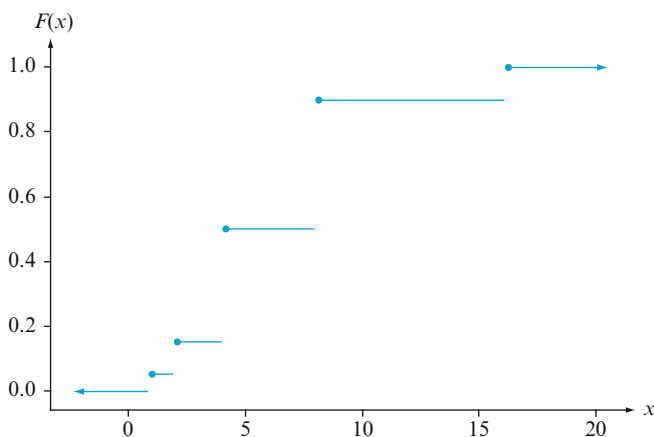

Fig. 2.5 A graph of the cdf of Example 2.11   ∎

For X a discrete rv, the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a **step function**.

Example 2.12 In Example 2.10, any positive integer was a possible X value, and the pmf was

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

For any positive integer x,

$$F(x) = \sum_{y \leq x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1}p = p\sum_{y=0}^{x-1} (1-p)^{y} \tag{2.4}$$

To evaluate this sum, we use the fact that the partial sum of a geometric series is

$$\sum_{y=0}^{k} a^{y} = \frac{1 - a^{k+1}}{1-a}$$

Using this in Eq. (2.4), with $a = 1 - p$ and $k = x - 1$, gives

$$F(x) = p \cdot \frac{1 - (1-p)^{x}}{1 - (1-p)} = 1 - (1-p)^{x} \qquad x \text{ a positive integer}$$

Since F is constant in between positive integers,

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - (1-p)^{[x]} & x \geq 1 \end{cases} \tag{2.5}$$

where $[x]$ is the largest integer $\leq x$ (e.g., $[2.7] = 2$). Thus if $p = .51$ as in the birth example, then the probability of having to examine at most five births to see the first boy is $F(5) = 1 - (.49)^5 = 1 - .0282 = .9718$, whereas $F(10) \approx 1.0000$. This cdf is graphed in Fig. 2.6.
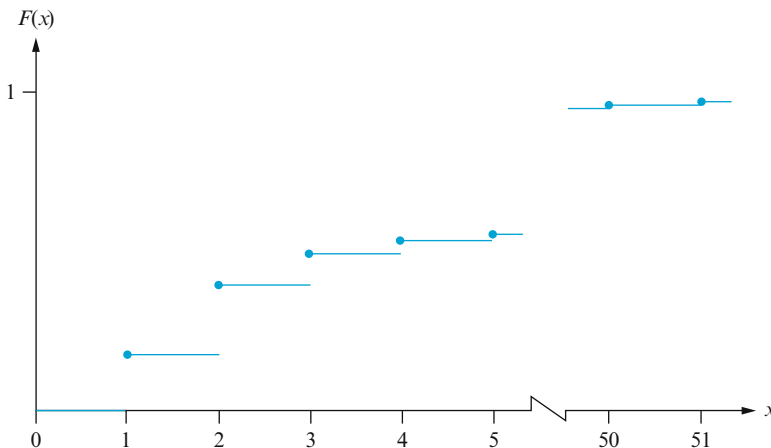

Fig. 2.6 A graph of $F(x)$ for Example 2.12 ∎

In our examples thus far, the cdf has been derived from the pmf. This process can be reversed to obtain the pmf from the cdf whenever the latter function is available. Suppose, for example, that X represents the number of defective components in a shipment consisting of six components, so that possible X values are 0, 1, ..., 6. Then

$$
\begin{aligned}
p(3) &= P(X = 3) \\
&= [p(0) + p(1) + p(2) + p(3)] - [p(0) + p(1) + p(2)] \\
&= P(X \leq 3) - P(X \leq 2) \\
&= F(3) - F(2)
\end{aligned}
$$

More generally, the probability that X falls in a specified interval is easily obtained from the cdf. For example,

$$
\begin{aligned}
P(2 \leq X \leq 4) &= p(2) + p(3) + p(4) \\
&= [p(0) + \cdots + p(4)] - [p(0) + p(1)] \\
&= P(X \leq 4) - P(X \leq 1) \\
&= F(4) - F(1)
\end{aligned}
$$

Notice that $P(2 \leq X \leq 4) \neq F(4) - F(2)$. This is because the X value 2 is included in $2 \leq X \leq 4$, so we do not want to subtract out its probability. However, $P(2 < X \leq 4) = F(4) - F(2)$ because $X = 2$ is not included in the interval $2 < X \leq 4$.

PROPOSITION

For any two numbers a and b with $a \leq b$,

$$
P(a \leq X \leq b) = F(b) - F(a-)
$$

where "$a-$" represents the largest possible X value that is strictly less than a. In particular, if the only possible values are integers and if a and b are integers, then

$$
\begin{aligned}
P(a \leq X \leq b) &= P(X = a \text{ or } a + 1 \text{ or} \ldots \text{or } b) \\
&= F(b) - F(a - 1)
\end{aligned}
$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a - 1)$ in this case.

The reason for subtracting $F(a-)$ rather than $F(a)$ is that we want to include $P(X = a)$; $F(b) - F(a)$ gives $P(a < X \leq b)$. This proposition will be used extensively when computing binomial and Poisson probabilities in Sects. 2.4 and 2.5.

Example 2.13 Let $X =$ the number of days of sick leave taken by a randomly selected employee of a large company during a particular year. If the maximum number of allowable sick days per year is 14, possible values of X are 0, 1, ..., 14. With $F(0) = .58, F(1) = .72, F(2) = .76, F(3) = .81, F(4) = .88$, and $F(5) = .94$,

$$
P(2 \leq X \leq 5) = P(X = 2, 3, 4, \text{ or } 5) = F(5) - F(1) = .22
$$

and

$$
P(X = 3) = F(3) - F(2) = .05 \qquad\blacksquare
$$

2.2.3 Another View of Probability Mass Functions

It is often helpful to think of a pmf as specifying a mathematical model for a discrete population.

Example 2.14   Consider selecting at random a household in a certain region, and Let $X =$ the number of individuals in the selected household. Suppose the pmf of X is as follows:

x	1	2	3	4	5	6	7	8	9	10
$p(x)$	.140	.175	.220	.260	.155	.025	.015	.005	.004	.001

This is very close to the household size distribution for rural Thailand given in the article "The Probability of Containment for Multitype Branching Process Models for Emerging Epidemics" (*J. of Applied Probability*, 2011: 173–188), which modeled influenza transmission.

Suppose this is based on one million households. One way to view this situation is to think of the population as consisting of 1,000,000 households, each one having its own X value; the proportion with each X value is given by $p(x)$ in the above table. An alternative viewpoint is to forget about the households and think of the population itself as consisting of X values—14% of these values are 1, 17.5% are 2, and so on. The pmf then describes the distribution of the possible population values 1, 2, …, 10. ∎

Once we have such a population model, we will use it to compute values of various population characteristics such as the *mean*, which describes the center of the population distribution, and the *standard deviation*, which describes the extent of spread about the center. Both of these are developed in the next section.

2.2.4   Exercises: Section 2.2 (11–28)

11. Let X be the number of students who show up at a professor's office hours on a particular day. Suppose that the only possible values of X are 0, 1, 2, 3, and 4, and that $p(0) = .30$, $p(1) = .25$, $p(2) = .20$, and $p(3) = .15$.
    (a) What is $p(4)$?
    (b) Draw both a line graph and a probability histogram for the pmf of X.
    (c) What is the probability that at least two students come to the office hour? What is the probability that more than two students come to the office hour?
    (d) What is the probability that the professor shows up for his office hour?
12. Airlines sometimes overbook flights. Suppose that for a plane with 50 seats, 55 passengers have tickets. Define the random variable Y as the number of ticketed passengers who actually show up for the flight. The probability mass function of Y appears in the accompanying table.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	.05	.10	.12	.14	.25	.17	.06	.05	.03	.02	.01

    (a) What is the probability that the flight will accommodate all ticketed passengers who show up?
    (b) What is the probability that not all ticketed passengers who show up can be accommodated?
    (c) If you are the first person on the standby list (which means you will be the first one to get on the plane if there are any seats available after all ticketed passengers have been accommodated), what is the probability that you will be able to take the flight? What is this probability if you are the third person on the standby list?

13. A mail-order computer business has six telephone lines. Let X denote the number of lines in use at a specified time. Suppose the pmf of X is as given in the accompanying table.

x	0	1	2	3	4	5	6
$p(x)$	.10	.15	.20	.25	.20	.06	.04

Calculate the probability of each of the following events.
(a) {at most three lines are in use}
(b) {fewer than three lines are in use}
(c) {at least three lines are in use}
(d) {between two and five lines, inclusive, are in use}
(e) {between two and four lines, inclusive, are not in use}
(f) {at least four lines are not in use}

14. A contractor is required by a county planning department to submit one, two, three, four, or five forms (depending on the nature of the project) in applying for a building permit. Let $Y =$ the number of forms required of the next applicant. The probability that y forms are required is known to be proportional to y—that is, $p(y) = ky$ for $y = 1, \ldots, 5$.
(a) What is the value of k? [*Hint:* $\sum_{y=1}^{5} p(y) = 1.$]
(b) What is the probability that at most three forms are required?
(c) What is the probability that between two and four forms (inclusive) are required?
(d) Could $p(y) = y^2/50$ for $y = 1, \ldots, 5$ be the pmf of Y?

15. Many manufacturers have quality control programs that include inspection of incoming materials for defects. Suppose a computer manufacturer receives computer boards in lots of five. Two boards are selected from each lot for inspection. We can represent possible outcomes of the selection process by pairs. For example, the pair (1, 2) represents the selection of boards 1 and 2 for inspection.
(a) List the ten different possible outcomes.
(b) Suppose that boards 1 and 2 are the only defective boards in a lot of five. Two boards are to be chosen at random. Define X to be the number of defective boards observed among those inspected. Find the probability distribution of X.
(c) Let $F(x)$ denote the cdf of X. First determine $F(0) = P(X \le 0)$, $F(1)$, and $F(2)$, and then obtain $F(x)$ for all other x.

16. Some parts of California are particularly earthquake-prone. Suppose that in one such area, 25% of all homeowners are insured against earthquake damage. Four homeowners are to be selected at random; let X denote the number among the four who have earthquake insurance.
(a) Find the probability distribution of X. [*Hint:* Let S denote a homeowner who has insurance and F one who does not. Then one possible outcome is $SFSS$, with probability (.25)(.75)(.25)(.25) and associated X value 3. There are 15 other outcomes.]
(b) Draw the corresponding probability histogram.
(c) What is the most likely value for X?
(d) What is the probability that at least two of the four selected have earthquake insurance?

17. A new battery's voltage may be acceptable (A) or unacceptable (U). A certain flashlight requires two batteries, so batteries will be independently selected and tested until two acceptable ones have been found. Suppose that 90% of all batteries have acceptable voltages. Let Y denote the number of batteries that must be tested.
(a) What is $p(2)$, that is, $P(Y = 2)$?
(b) What is $p(3)$? [*Hint:* There are two different outcomes that result in $Y = 3$.]

(c) To have $Y = 5$, what must be true of the fifth battery selected? List the four outcomes for which $Y = 5$ and then determine $p(5)$.

(d) Use the pattern in your answers for parts (a)–(c) to obtain a general formula for $p(y)$.

18. Two fair six-sided dice are tossed independently. Let $M =$ the maximum of the two tosses, so $M(1, 5) = 5$, $M(3, 3) = 3$, etc.

(a) What is the pmf of M? [*Hint:* First determine $p(1)$, then $p(2)$, and so on.]

(b) Determine the cdf of M and graph it.

19. A library subscribes to two different weekly news magazines, each of which is supposed to arrive in Wednesday's mail. In actuality, each one may arrive on Wednesday, Thursday, Friday, or Saturday. Suppose the two arrive independently of one another, and for each one $P(W) = .3$, $P(\text{Th}) = .4$, $P(F) = .2$, and $P(S) = .1$. Let $Y =$ the number of days beyond Wednesday that it takes for both magazines to arrive (so possible Y values are 0, 1, 2, or 3). Compute the pmf of Y. [*Hint:* There are 16 possible outcomes; $Y(W, W) = 0$, $Y(F, \text{Th}) = 2$, and so on.]

20. Three couples and two single individuals have been invited to an investment seminar and have agreed to attend. Suppose the probability that any particular couple or individual arrives late is .4 (a couple will travel together in the same vehicle, so either both people will be on time or else both will arrive late). Assume that different couples and individuals are on time or late independently of one another. Let $X =$ the number of people who arrive late for the seminar.

(a) Determine the probability mass function of X. [*Hint:* label the three couples #1, #2, and #3 and the two individuals #4 and #5.]

(b) Obtain the cumulative distribution function of X, and use it to calculate $P(2 \leq X \leq 6)$.

21. As described in the book's Introduction, *Benford's Law* arises in a variety of situations as a model for the first digit of a number:

$$p(x) = P(\text{1st digit is } x) = \log_{10}\left(\frac{x + 1}{x}\right), x = 1, 2, \ldots, 9$$

(a) Without computing individual probabilities from this formula, show that it specifies a legitimate pmf.

(b) Now compute the individual probabilities and compare to the distribution where 1, 2, ..., 9 are equally likely.

(c) Obtain the cdf of X, a rv following Benford's law.

(d) Using the cdf, what is the probability that the leading digit is at most 3? At least 5?

22. Refer to Exercise 13, and calculate and graph the cdf $F(x)$. Then use it to calculate the probabilities of the events given in parts (a)–(d) of that problem.

23. Let X denote the number of vehicles queued up at a bank's drive-up window at a particular time of day. The cdf of X is as follows:

$$F(x) = \begin{cases} 0 & x < 0 \\ .06 & 0 \leq x < 1 \\ .19 & 1 \leq x < 2 \\ .39 & 2 \leq x < 3 \\ .67 & 3 \leq x < 4 \\ .92 & 4 \leq x < 5 \\ .97 & 5 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$

Calculate the following probabilities directly from the cdf:

(a) $p(2)$, that is, $P(X = 2)$

(b) $P(X > 3)$

(c) $P(2 \leq X \leq 5)$

(d) $P(2 < X < 5)$

24. An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let $X =$ the number of months between successive payments. The cdf of X is as follows:

$$F(x) = \begin{cases} 0 & x < 1 \\ .30 & 1 \leq x < 3 \\ .40 & 3 \leq x < 4 \\ .45 & 4 \leq x < 6 \\ .60 & 6 \leq x < 12 \\ 1 & 12 \leq x \end{cases}$$

(a) What is the pmf of X?

(b) Using just the cdf, compute $P(3 \leq X \leq 6)$ and $P(4 \leq X)$.

25. In Example 2.10, let $Y =$ the number of girls born before the experiment terminates. With $p = P(B)$ and $1 - p = P(G)$, what is the pmf of Y? [*Hint:* First list the possible values of Y, starting with the smallest, and proceed until you see a general formula.]

26. Alvie Singer lives at 0 in the accompanying diagram and has four friends who live at A, B, C, and D. One day Alvie decides to go visiting, so he tosses a fair coin twice to decide which of the four to visit. Once at a friend's house, he will either return home or else proceed to one of the two adjacent houses (such as 0, A, or C when at B), with each of the three possibilities having probability 1/3. In this way, Alvie continues to visit friends until he returns home.
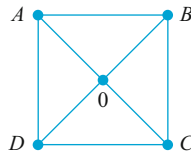
(a) Let $X =$ the number of times that Alvie visits a friend. Derive the pmf of X.

(b) Let $Y =$ the number of straight-line segments that Alvie traverses (including those leading to and from 0). What is the pmf of Y?

(c) Suppose that female friends live at A and C and male friends at B and D. If $Z =$ the number of visits to female friends, what is the pmf of Z?

27. After all students have left the classroom, a statistics professor notices that four copies of the text were left under desks. At the beginning of the next lecture, the professor distributes the four books in a completely random fashion to each of the four students (1, 2, 3, and 4) who claim to have left books. One possible outcome is that 1 receives 2's book, 2 receives 4's book, 3 receives his or her own book, and 4 receives 1's book. This outcome can be abbreviated as (2, 4, 3, 1).

(a) List the other 23 possible outcomes.

(b) Let X denote the number of students who receive their own book. Determine the pmf of X.

28. Show that the cdf $F(x)$ is a nondecreasing function; that is, $x_1 < x_2$ implies that $F(x_1) \leq F(x_2)$. Under what condition will $F(x_1) = F(x_2)$?

2.3   **Expected Value and Standard Deviation**

Consider a university with 15,000 students and let $X =$ the number of courses for which a randomly selected student is registered. The pmf of X follows. Since $p(1) = .01$, we know that $(.01) \cdot (15{,}000) = 150$ of the students are registered for one course, and similarly for the other x values.

x	1	2	3	4	5	6	7	
$p(x)$	.01	.03	.13	.25	.39	.17	.02	(2.6)
Number registered	150	450	1950	3750	5850	2550	300	

To compute the average number of courses per student, i.e., the average value of X in the population, we should calculate the total number of courses and divide by the total number of students. Since each of 150 students is taking one course, these 150 contribute 150 courses to the total. Similarly, 450 students contribute 2(450) courses, and so on. The population average value of X is then

$$\frac{1(150) + 2(450) + 3(1950) + \cdots + 7(300)}{15{,}000} = 4.57 \tag{2.7}$$

Since $150/15{,}000 = .01 = p(1)$, $450/15{,}000 = .03 = p(2)$, and so on, an alternative expression for Eq. (2.7) is

$$1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \tag{2.8}$$

Expression (2.8) shows that to compute the population average value of X, we need only the possible values of X along with their probabilities (proportions). In particular, the population size is irrelevant as long as the pmf is given by (2.6). The average or mean value of X is then a *weighted* average of the possible values 1, …, 7, where the weights are the probabilities of those values.

2.3.1   **The Expected Value of X**

> **DEFINITION**
> Let X be a discrete rv with set of possible values D and pmf $p(x)$. The **expected value** or **mean value** of X, denoted by $E(X)$ or μ_X or just μ, is
>
> $$E(X) = \mu_X = \mu = \sum_{x \in D} x \cdot p(x)$$

Example 2.15   For the pmf of $X =$ number of courses in (2.6),

$$\begin{aligned}
\mu &= 1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \\
&= (1)(.01) + (2)(.03) + \cdots + (7)(.02) \\
&= .01 + .06 + .39 + 1.00 + 1.95 + 1.02 + .14 = 4.57
\end{aligned}$$

If we think of the population as consisting of the X values 1, 2, …, 7, then $\mu = 4.57$ is the population mean (we will often refer to μ as the *population mean* rather than the mean of X in the population). Notice that μ here is not 4, the ordinary average of 1, …, 7, because the distribution puts more weight on 4, 5, and 6 than on other X values. ∎

In Example 2.15, the expected value μ was 4.57, which is not a possible value of X. The word *expected* should be interpreted with caution because one would not expect to see an X value of 4.57 when a single student is selected.

Example 2.16 Just after birth, each newborn child is rated on a scale called the Apgar scale. The possible ratings are 0, 1, …, 10, with the child's rating determined by color, muscle tone, respiratory effort, heartbeat, and reflex irritability (the best possible score is 10). Let X be the Apgar score of a randomly selected child born at a certain hospital during the next year, and suppose that the pmf of X is

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

Then the mean value of X is

$$E(X) = \mu = (0)(.002) + (1)(.001) + (2)(.002) + \cdots + (8)(.25) + (9)(.12) + (10)(.01)$$
$$= 7.15$$

(Again, μ is not a possible value of the variable X.) If the stated model is correct, then the mean Apgar score for the population of all children born at this hospital next year will be 7.15. ∎

Example 2.17 Let $X = 1$ if a randomly selected component needs warranty service and $= 0$ otherwise. If the chance a component needs warranty service is p, then X is a Bernoulli rv with pmf $p(1) = p$ and $p(0) = 1 - p$, from which

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1 - p) + 1(p) = p$$

That is, the expected value of X is just the probability that X takes on the value 1. If we conceptualize a population consisting of 0s in proportion $1 - p$ and 1s in proportion p, then the population average is $\mu = p$. ∎

There is another frequently used interpretation of μ. Consider observing a first value x_1 of X, then a second value x_2, a third value x_3, and so on. After doing this a large number of times, calculate the sample average of the observed x_is. This average will typically be close to μ; a more rigorous version of this statement is provided by the *Law of Large Numbers* in Chap. 4. That is, μ can be interpreted as the long-run average value of X when the experiment is performed repeatedly. This interpretation is often appropriate for games of chance, where the "population" is not a concrete set of individuals but rather the results of all hypothetical future instances of playing the game.

Example 2.18 A standard American roulette wheel has 38 spaces. Players bet on which space a marble will land in once the wheel has been spun. One of the simplest bets is based on the color of the space: 18 spaces are black, 18 are red, and 2 are green. So, if a player "bets on black," s/he has an 18/38 chance of winning. Casinos consider color bets an "even wager," meaning that a player who bets $1 on black, say, will profit $1 if the marble lands in a black space (and lose the wagered $1 otherwise).

Let $X = $ the return on a $1 wager on black. Then the pmf of X is

x	−$1	+$1
$p(x)$	20/38	18/38

and the expected value of X is $E(X) = (-1)(20/38) + (1)(18/38) = -2/38 = -\$.0526$. If a player makes \$1 bets on black on successive spins of the roulette wheel, in the long run s/he can expect to lose about 5.26 cents per wager. Since players don't necessarily make a large number of wagers, this long-run average interpretation is perhaps more apt from the casino's perspective: in the long run, they will gain an average of 5.26 cents for every \$1 wagered on black at the roulette table. ∎

Thus far, we have assumed that the mean of any given distribution exists. If the set of possible values of X is unbounded, so that the sum for μ_X is actually an infinite series, the expected value of X might or might not exist (depending on whether the series converges or diverges).

Example 2.19   From Example 2.10, the general form for the pmf of $X = $ the number of children born up to and including the first boy is

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

The expected value of X therefore entails evaluating an infinite summation:

$$E(X) = \sum_D x \cdot p(x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = p\sum_{x=1}^{\infty} x(1-p)^{x-1} = p\sum_{x=1}^{\infty} \left[-\frac{d}{dp}(1-p)^x \right] \qquad (2.9)$$

If we interchange the order of taking the derivative and the summation in Eq. (2.9), the sum is that of a geometric series. (In particular, the infinite series converges for $0 < p < 1$.)

After the sum is computed and the derivative is taken, the final result is $E(X) = 1/p$. That is, the expected number of children born up to and including the first boy is the reciprocal of the chance of getting a boy. This is actually quite intuitive: if p is near 1, we expect to see a boy very soon, whereas if p is near 0, we expect many births before the first boy. For $p = .5$, $E(X) = 2$.

Exercise 48 at the end of this section presents an alternative method for computing the mean of this particular distribution. ∎

Example 2.20   Let X, the number of interviews a student has prior to getting a job, have pmf

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where k is such that $\sum_{x=1}^{\infty}(k/x^2) = 1$. (Because $\sum_{x=1}^{\infty}(1/x^2) = \pi^2/6$, the value of k is $6/\pi^2$.) The expected value of X is

$$\mu = E(X) = \sum_{x=1}^{\infty} x\frac{k}{x^2} = k\sum_{x=1}^{\infty} \frac{1}{x} \qquad (2.10)$$

The sum on the right of Eq. (2.10) is the famous *harmonic series* of mathematics and can be shown to diverge. $E(X)$ is not finite here because $p(x)$ does not decrease sufficiently fast as x increases; statisticians say that the probability distribution of X has "a heavy tail." If a sequence of X values is chosen using this distribution, the sample average will not settle down to some finite number but will tend to grow without bound. ∎

2.3.2 The Expected Value of a Function

Often we will be interested in the expected value of some function $h(X)$ rather than X itself. An easy way of computing the expected value of $h(X)$ is suggested by the following example.

Example 2.21 The cost of a certain vehicle diagnostic test depends on the number of cylinders X in the vehicle's engine. Suppose the cost function is $h(X) = 20 + 3X + .5X^2$. Since X is a random variable, so is $Y = h(X)$. The pmf of X and the derived pmf of Y are as follows:

x	4	6	8
$p(x)$	.5	.3	.2

$\Rightarrow$

y	40	56	76
$p(y)$	.5	.3	.2

With D^* denoting possible values of Y,

$$E(Y) = E[h(X)] = \sum_{y \in D^*} y \cdot p(y) = (40)(.5) + (56)(.3) + (76)(.2) = \$52$$

$$= h(4) \cdot (.5) + h(6) \cdot (.3) + h(8) \cdot (.2) = \sum_D h(x) \cdot p(x) \tag{2.11}$$

According to Eq. (2.11), it was not necessary to determine the pmf of Y to obtain $E(Y)$; instead, the desired expected value is a weighted average of the possible $h(x)$ (rather than x) values. ∎

PROPOSITION

If the rv X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

This is sometimes referred to as the *Law of the Unconscious Statistician.*

According to this proposition, $E[h(X)]$ is computed in the same way that $E(X)$ itself is, except that $h(x)$ is substituted in place of x. That is, $E[h(X)]$ is a weighted average of possible $h(X)$ values, where the weights are the probabilities of the corresponding original X values.

Example 2.22 A computer store has purchased three computers at \$500 apiece. It will sell them for \$1,000 apiece. The manufacturer has agreed to repurchase any computers still unsold after a specified period at \$200 apiece. Let X denote the number of computers sold, and suppose that $p(0) = .1$, $p(1) = .2$, $p(2) = .3$, and $p(3) = .4$. With $h(X)$ denoting the profit associated with selling X units, the given information implies that $h(X) = \text{revenue} - \text{cost} = 1000X + 200(3 - X) - 1500 = 800X - 900$. The expected profit is then

$$\begin{aligned}
E[h(X)] &= h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) \\
&= (800(0) - 900)(.1) + (800(1) - 900)(.2) + (800(2) - 900)(.3) + (800(3) - 900)(.4) \\
&= (-900)(.1) + (-100)(.2) + (700)(.3) + (1500)(.4) = \$700 \qquad\blacksquare
\end{aligned}$$

Because an expected value is a sum, it possesses the same properties as any summation; specifically, the expected value "operator" can be distributed across addition and across multiplication by constants. This important property is known as *linearity of expectation*.

LINEARITY OF EXPECTATION

For any functions $h_1(X)$ and $h_2(X)$ and any constants a_1, a_2, and b,

$$E[a_1 h_1(X) + a_2 h_2(X) + b] = a_1 E[h_1(X)] + a_2 E[h_2(X)] + b$$

In particular, for any linear function $aX + b$,

$$E(aX + b) = a \cdot E(X) + b \tag{2.12}$$

(or, using alternative notation, $\mu_{aX+b} = a \cdot \mu_X + b$).

Proof Let $h(X) = a_1 h_1(X) + a_2 h_2(X) + b$, and apply the previous proposition:

$$E[a_1 h_1(X) + a_2 h_2(X) + b] = \sum_D (a_1 h_1(x) + a_2 h_2(x) + b) \cdot p(x)$$

$$= a_1 \sum_D h_1(x) \cdot p(x) + a_2 \sum_D h_2(x) \cdot p(x)$$

$$+ b \sum_D p(x) \qquad \text{distributive property of addition}$$

$$= a_1 E[h_1(X)] + a_2 E[h_2(X)] + b[1] = a_1 E[h_1(X)] + a_2 E[h_2(X)] + b$$

The special case of $aX + b$ is obtained by setting $a_1 = a$, $h_1(X) = X$, and $a_2 = 0$. ∎

By induction, linearity of expectation applies to any finite number of terms. In Example 2.21, it is easily computed that $E(X) = 4(.5) + 6(.3) + 8(.2) = 5.4$ and $E(X^2) = \sum x^2 \cdot p(x) = 4^2(.5) + 6^2(.3) + 8^2(.2) = 31.6$. Applying linearity of expectation to $Y = h(X) = 20 + 3X + .5X^2$, we obtain

$$\mu_Y = E[20 + 3X + .5X^2] = 20 + 3E(X) + .5E(X^2) = 20 + 3(5.4) + .5(31.6) = \$52,$$

which matches the result of Example 2.21.

The special case Eq. (2.12) states that the expected value of a linear function equals the linear function evaluated at the expected value $E(X)$. Since $h(X)$ in Example 2.22 is linear and $E(X) = 2$, $E[h(X)] = 800(2) - 900 = \$700$, as before. Two special cases of Eq. (2.12) yield two important rules of expected value.

1. For any constant a, $\mu_{aX} = a \cdot \mu_X$ (take $b = 0$).
2. For any constant b, $\mu_{X+b} = \mu_X + b = E(X) + b$ (take $a = 1$).

Multiplication of X by a constant a changes the unit of measurement (from dollars to cents, where $a = 100$, inches to cm, where $a = 2.54$, etc.). Rule 1 says that the expected value in the new units equals the expected value in the old units multiplied by the conversion factor a. Similarly, if the constant b is added to each possible value of X, then the expected value will be shifted by that same amount.

One commonly made error is to substitute μ_X directly into the function $h(X)$ when h is a nonlinear function, in which case Eq. (2.12) does not apply. Consider Example 2.21: the mean of X is 5.4, and it's tempting to infer that the mean of $Y = h(X)$ is simply $h(5.4)$. However, since the function $h(X) = 20 + 3X + .5X^2$ is *not* linear, this does not yield the correct answer:

$$h(5.4) = 20 + 3(5.4) + .5(5.4)^2 = 50.78 \neq 52 = \mu_Y$$

In general, $\mu_{h(X)}$ *does not equal* $h(\mu_X)$ *unless the function* $h(x)$ *is linear.*

2.3.3   The Variance and Standard Deviation of *X*

The expected value of *X* describes where the probability distribution is centered. Using the physical analogy of placing point mass $p(x)$ at the value x on a one-dimensional axis, if the axis were then supported by a fulcrum placed at μ, there would be no tendency for the axis to tilt. This is illustrated for two different distributions in Fig. 2.7.
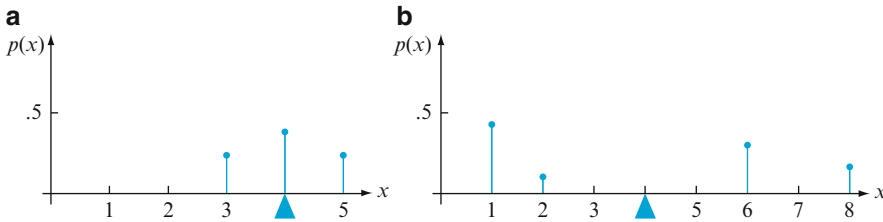

Fig. 2.7   Two different probability distributions with $\mu = 4$

Although both distributions pictured in Fig. 2.7 have the same mean/fulcrum μ, the distribution of Fig. 2.7b has greater spread or variability or dispersion than does that of Fig. 2.7a. Our goal now is to obtain a quantitative assessment of the extent to which the distribution spreads out about its mean value.

> **DEFINITION**
> Let *X* have pmf $p(x)$ and expected value μ. Then the **variance** of *X*, denoted by $\mathrm{Var}(X)$ or σ_X^2 or just σ^2, is
>
> $$\mathrm{Var}(X) = \sum_D \left[(x - \mu)^2 \cdot p(x) \right] = E\left[(X - \mu)^2 \right]$$
>
> The **standard deviation** (SD) of *X*, denoted by $\mathrm{SD}(X)$ or σ_X or just σ, is
>
> $$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

The quantity $h(X) = (X - \mu)^2$ is the squared deviation of *X* from its mean, and σ^2 is the expected squared deviation—i.e., a weighted average of the squared deviations from μ. Taking the square root of the variance to obtain standard deviation returns us to the original units of the variable, e.g., if *X* is measured in dollars, then both μ and σ also have units of dollars. If most of the probability distribution is close to μ, as in Fig. 2.7a, then σ will typically be relatively small. However, if there are *x* values far from μ that have large probabilities (as in Fig. 2.7b), then σ will be larger.

Example 2.23   Consider again the distribution of the Apgar score *X* of a randomly selected newborn described in Example 2.16. The mean value of *X* was calculated as $\mu = 7.15$, so

$$\mathrm{Var}(X) = \sigma^2 = \sum_{x=0}^{10} (x - 7.15)^2 \cdot p(x) = (0 - 7.15)^2(.002) + \ldots + (10 - 7.15)^2(.01) = 1.5815$$

The standard deviation of *X* is $\mathrm{SD}(X) = \sigma = \sqrt{1.5815} = 1.26$.   ■

A rough interpretation of σ is that its value gives the size of a typical or representative distance from μ (hence, "standard deviation"). Because $\sigma = 1.26$ in the preceding example, we can say that some of the possible X values differ by more than 1.26 from the mean value 7.15 whereas other possible X values are closer than this to 7.15; roughly, 1.26 is the size of a typical deviation from the mean Apgar score.

Example 2.24   (Example 2.18 continued) The variance of $X =$ the return on a \$1 bet on black is

$$\sigma_X^2 = (-1 - (-2/38))^2 \cdot (20/38) + (1 - (-2/38))^2 \cdot 18/38 = 0.99723$$

and the standard deviation is $\sigma_X = \sqrt{0.99723} = 0.9986 \approx \$1$. The two possible values of X are $-\$1$ and $+\$1$; since betting on black is almost a break-even wager (the mean is quite close to 0), the typical difference between an actual return X and the average return μ_X is roughly one dollar.   ∎

A natural probability question arises: how often does X fall within this "typical distance of the mean"? That is, what's the chance that a rv X lies between $\mu_X - \sigma_X$ and $\mu_X + \sigma_X$? What about the likelihood that X is within two standard deviations of its mean? There are no universal answers: for different pmfs, varying amounts of probability may lie within one (or two or three) standard deviation(s) of the expected value. That said, the following theorem, due to Russian mathematician Pafnuty Chebyshev, partially addresses questions of this sort.

> **CHEBYSHEV'S INEQUALITY**
> Let X be a discrete rv with mean μ and standard deviation σ. Then, for any $k \geq 1$,
>
> $$P\big(|X - \mu| \geq k\sigma\big) \leq \frac{1}{k^2}$$
>
> That is, the probability X is at least k standard deviations away from its mean is at most $1/k^2$.

An equivalent statement to Chebyshev's inequality is that every random variable has a probability of at least $1 - 1/k^2$ to fall within k standard deviations of its mean.

Proof   Let A denote the event $|X - \mu| \geq k\sigma$; or, equivalently, the set of values $\{x : |x - \mu| \geq k\sigma\}$. Begin by writing out the definition of $\mathrm{Var}(X)$:

$$\mathrm{Var}(X) = \sum_D \Big[(x - \mu)^2 \cdot p(x)\Big] = \sum_A \Big[(x - \mu)^2 \cdot p(x)\Big] + \sum_{A'} \Big[(x - \mu)^2 \cdot p(x)\Big]$$

$$\geq \sum_A \Big[(x - \mu)^2 \cdot p(x)\Big] \qquad \text{because the discarded term is} \geq 0$$

$$\geq \sum_A \Big[(k\sigma)^2 \cdot p(x)\Big] \qquad \text{because } (x - \mu)^2 \geq (k\sigma)^2 \text{ on the set } A$$

$$= (k\sigma)^2 \sum_A p(x) = (k\sigma)^2 P(A) = k^2 \sigma^2 P\big(|X - \mu| \geq k\sigma\big)$$

The $\mathrm{Var}(X)$ term on the left-hand side is the same as the σ^2 term on the right-hand side; cancelling the two, we are left with $1 \geq k^2 P(|X - \mu| \geq k\sigma)$, and Chebyshev's inequality follows.   ∎

For $k = 1$, Chebyshev's inequality states that $P(|X - \mu| \geq \sigma) \leq 1$, which isn't very informative since all probabilities are bounded above by 1. In fact, distributions can be constructed for which 100% of the distribution is at least 1 standard deviation from the mean, so that the rv X has probability 0 of falling less than one standard deviation from its mean (see Exercise 47). Substituting $k = 2$, Chebyshev's inequality states that the chance any rv is at least 2 standard deviations from its mean cannot exceed $1/2^2 = .25 = 25\%$. Equivalently, *every* distribution has the property that at least 75% of its "mass" lies within 2 standard deviations of its mean value (in fact, for many distributions, the exact probability is much larger than this lower bound).

2.3.4 Properties of Variance

An alternative to the defining formula for $\mathrm{Var}(X)$ reduces the computational burden.

PROPOSITION

$$\mathrm{Var}(X) = \sigma^2 = E(X^2) - \mu^2$$

This equation is referred to as the *variance shortcut formula*.

In using this formula, $E(X^2)$ is computed first without any subtraction; then μ is computed, squared, and subtracted (once) from $E(X^2)$. This formula is more efficient because it entails only one subtraction, and $E(X^2)$ does not require calculating squared deviations from μ.

Example 2.25 Referring back to the Apgar score scenario of Examples 2.16 and 2.23,

$$E(X^2) = \sum_{x=1}^{10} x^2 \cdot p(x) = (0^2)(.002) + (1^2)(.001) + \cdots + (10^2)(.01) = 52.704$$

Thus, $\sigma^2 = 52.704 - (7.15)^2 = 1.5815$ as before, and again $\sigma = 1.26$. ∎

Proof of the Variance Shortcut Formula Expand $(X - \mu)^2$ in the definition of $\mathrm{Var}(X)$, and then apply linearity of expectation:

$$
\begin{aligned}
\mathrm{Var}(X) = E\big[(X - \mu)^2\big] &= E\big[X^2 - 2\mu X + \mu^2\big] \\
&= E(X^2) - 2\mu E(X) + \mu^2 \qquad \text{by linearity of expectation} \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2
\end{aligned}
$$
∎

The quantity $E(X^2)$ in the variance shortcut formula is called the **mean-square value** of the random variable X. Engineers may be familiar with the root-mean-square, or RMS, which is the square root of $E(X^2)$. Do not confuse this with the square of the mean of X, i.e., μ^2! For example, if X has a mean of 7.15, the mean-square value of X is *not* $(7.15)^2$, because $h(x) = x^2$ is not linear. (In Example 2.25, the mean-square value of X is 52.704.) It helps to look at the two formulas side-by-side:

$$E(X^2) = \sum_D x^2 \cdot p(x) \quad \text{versus} \quad \mu^2 = \left(\sum_D x \cdot p(x) \right)^2$$

The order of operations is clearly different. In fact, it can be shown (see Exercise 46) that $E(X^2) \geq \mu^2$ for every random variable, with equality if and only if X is constant.

The variance of a function $h(X)$ is the expected value of the squared difference between $h(X)$ and its expected value:

$$\text{Var}[h(X)] = \sigma^2_{h(X)} = \sum_D \left[\left(h(x) - \mu_{h(X)} \right)^2 \cdot p(x) \right] = \left[\sum_D h^2(x) \cdot p(x) \right] - \left[\sum_D h(x) \cdot p(x) \right]^2$$

When $h(x)$ is a linear function, $\text{Var}[h(X)]$ has a much simpler expression (see Exercise 43 for a proof).

> **PROPOSITION**
>
> $$\text{Var}(aX + b) = \sigma^2_{aX+b} = a^2 \cdot \sigma^2_X \quad \text{and} \quad \sigma_{aX+b} = |a| \cdot \sigma_X \qquad (2.13)$$
>
> In particular,
>
> $$\sigma_{aX} = |a| \cdot \sigma_X \quad \text{and} \quad \sigma_{X+b} = \sigma_X$$

The absolute value is necessary because a might be negative, yet a standard deviation cannot be. Usually multiplication by a corresponds to a change in the unit of measurement (e.g., kg to lb or dollars to euros); the sd in the new unit is just the original sd multiplied by the conversion factor. On the other hand, the addition of the constant b does not affect the variance, which is intuitive, because the addition of b changes the location (mean value) but not the spread of values. Together, Eqs. (2.12) and (2.13) comprise the *rescaling properties* of mean and standard deviation.

Example 2.26 In the computer sales scenario of Example 2.22, $E(X) = 2$ and

$$E(X^2) = (0^2)(.1) + (1^2)(.2) + (2^2)(.3) + (3^2)(.4) = 5$$

so $\text{Var}(X) = 5 - (2)^2 = 1$. The profit function $Y = h(X) = 800X - 900$ is linear, so Eq. (2.13) applies with $a = 800$ and $b = -900$. Hence Y has variance $a^2 \sigma^2_X = (800)^2(1) = 640,000$ and standard deviation \$800. ∎

2.3.5   Exercises: Section 2.3 (29–48)

29. The pmf of the amount of memory X (GB) in a purchased flash drive was given in Example 2.11 as

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

   (a) Compute and interpret $E(X)$.
   (b) Compute $\text{Var}(X)$ directly from the definition.
   (c) Obtain and interpret the standard deviation of X.
   (d) Compute $\text{Var}(X)$ using the shortcut formula.

30. An individual who has automobile insurance from a company is randomly selected. Let Y be the number of moving violations for which the individual was cited during the last 3 years. The pmf of Y is

y	0	1	2	3
$p(y)$	.60	.25	.10	.05

(a) Compute $E(Y)$.
(b) Suppose an individual with Y violations incurs a surcharge of $\$100Y^2$. Calculate the expected amount of the surcharge.

31. Refer to Exercise 12 and calculate $\text{Var}(Y)$ and σ_Y. Then determine the probability that Y is within 1 standard deviation of its mean value.

32. An appliance dealer sells three different models of upright freezers having 13.5, 15.9, and 19.1 cubic feet of storage space, respectively. Let $X =$ the amount of storage space purchased by the next customer to buy a freezer. Suppose that X has pmf

x	13.5	15.9	19.1
$p(x)$	.2	.5	.3

(a) Compute $E(X)$, $E(X^2)$, and $\text{Var}(X)$.
(b) If the price of a freezer having capacity X cubic feet is $17X + 180$, what is the expected price paid by the next customer to buy a freezer?
(c) What is the standard deviation of the price $17X + 180$ paid by the next customer?
(d) Suppose that although the rated capacity of a freezer is X, the actual capacity is $h(X) = X - .01X^2$. What is the expected actual capacity of the freezer purchased by the next customer?

33. Let X be a Bernoulli rv with pmf as in Example 2.17.
(a) Compute $E(X^2)$.
(b) Show that $\text{Var}(X) = p(1 - p)$.
(c) Compute $E(X^{79})$.

34. Suppose that the number of plants of a particular type found in a rectangular sampling region (called a quadrat by ecologists) in a certain geographic area is an rv X with pmf

$$p(x) = \begin{cases} c/x^3 & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Is $E(X)$ finite? Justify your answer. (This is another distribution that statisticians would call heavy-tailed.)

35. A small market orders copies of a certain magazine for its magazine rack each week. Let $X =$ demand for the magazine, with pmf

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suppose the store owner actually pays $\$2.00$ for each copy of the magazine and the price to customers is $\$4.00$. If magazines left at the end of the week have no salvage value, is it better to order three or four copies of the magazine? [*Hint:* For both three and four copies ordered, express net revenue as a function of demand X, and then compute the expected revenue.]

36. Let X be the damage incurred (in $) in a certain type of accident during a given year. Possible X values are 0, 1000, 5000, and 10,000, with probabilities .8, .1, .08, and .02, respectively. A particular company offers a $500 deductible policy. If the company wishes its expected profit to be $100, what premium amount should it charge?

37. The n candidates for a job have been ranked 1, 2, 3, ..., n. Let $X =$ the rank of a randomly selected candidate, so that X has pmf

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

(this is called the *discrete uniform distribution*). Compute $E(X)$ and $\text{Var}(X)$ using the shortcut formula. [*Hint:* The sum of the first n positive integers is $n(n + 1)/2$, whereas the sum of their squares is $n(n + 1)(2n + 1)/6$.]

38. Let $X =$ the outcome when a fair die is rolled once. If before the die is rolled you are offered either $100 dollars or $h(X) = 350/X$ dollars, would you accept the guaranteed amount or would you gamble? [*Hint:* Determine $E[h(X)]$, but be careful: the mean of $350/X$ is not $350/\mu$.]

39. In the popular game Plinko on *The Price Is Right*, contestants drop a circular disk (a "chip") down a pegged board; the chip bounces down the board and lands in a slot corresponding to one of five dollar mounts. The random variable $X =$ winnings from one chip dropped from the middle slot has roughly the following distribution.

x	$0	$100	$500	$1000	$10,000
$p(x)$	.39	.03	.11	.24	.23

(a) Graph the probability mass function of X.
(b) What is the probability a contestant makes money on a chip?
(c) What is the probability a contestant makes at least $1000 on a chip?
(d) Determine the expected winnings. Interpret this number.
(e) Determine the corresponding standard deviation.

40. A supply company currently has in stock 500 lb of fertilizer, which it sells to customers in 10-lb bags. Let X equal the number of bags purchased by a randomly selected customer. Sales data shows that X has the following pmf:

x	1	2	3	4
$p(x)$	.2	.4	.3	.1

(a) Compute the average number of bags bought per customer.
(b) Determine the standard deviation for the number of bags bought per customer.
(c) Define Y to be the amount of fertilizer left in stock, in pounds, after the first customer. Construct the pmf of Y.
(d) Use the pmf of Y to find the expected amount of fertilizer left in stock, in pounds, after the first customer.
(e) Write Y as a linear function of X. Then use rescaling properties to find the mean and standard deviation of Y.
(f) The supply company offers a discount to each customer based on the formula $W = (X - 1)^2$. Determine the expected discount for a customer.
(g) Does your answer in part (f) equal $(\mu_X - 1)^2$? Why or why not?
(h) Calculate the standard deviation of W.

41. Refer back to the roulette scenario in Examples 2.18 and 2.24. Two other ways to wager at roulette are betting on a single number, or on a four-number "square." The pmfs for the returns on a \$1 wager on a number and a square are displayed below. (Payoffs for winning are always based on the odds of losing a wager under the assumption the two green spaces didn't exist.)
Single number:

x	$-$\$1	$+$\$35
$p(x)$	37/38	1/38

Square:

x	$-$\$1	$+$\$8
$p(x)$	34/38	4/38

(a) Determine the expected return from a \$1 wager on a single number, and then on a square.

(b) Compare your answers from (a) to Example 2.18. What can be said about the expected return for a \$1 wager? Based on this, does expected return reflect most players' intuition that betting on black is "safer" and betting on a single number is "riskier"?

(c) Now calculate the standard deviations for the two pmfs above.

(d) How do the standard deviations of the three betting schemes (color, single number, square) compare? How do these values appear to relate to players' intuitive sense of risk?

42. (a) Draw a line graph of the pmf of X in Exercise 35. Then determine the pmf of $-X$ and draw its line graph. From these two pictures, what can you say about $\text{Var}(X)$ and $\text{Var}(-X)$?

(b) Use the proposition involving $\text{Var}(aX + b)$ to establish a general relationship between $\text{Var}(X)$ and $\text{Var}(-X)$.

43. Use the definition of variance to prove that $\text{Var}(aX + b) = a^2 \sigma_X^2$. [*Hint:* From Eq. (2.12), $\mu_{aX+b} = a\mu_X + b$.]

44. Suppose $E(X) = 5$ and $E[X(X - 1)] = 27.5$.
(a) Determine $E(X^2)$. [*Hint:* $E[X(X - 1)] = E(X^2 - X) = E(X^2) - E(X)$.]
(b) What is $\text{Var}(X)$?
(c) What is the general relationship among the quantities $E(X)$, $E[X(X - 1)]$, and $\text{Var}(X)$?

45. Write a general rule for $E(X - c)$ where c is a constant. What happens when you let $c = \mu$, the expected value of X?

46. Let X be a rv with mean μ. Show that $E(X^2) \geq \mu^2$, and that $E(X^2) > \mu^2$ unless X is a constant. [*Hint:* Consider variance.]

47. Refer to Chebyshev's inequality in this section.
(a) What is the value of the upper bound for $k = 2$? $k = 3$? $k = 4$? $k = 5$? $k = 10$?
(b) Compute μ and σ for the distribution of Exercise 13. Then evaluate for the values of k given in part (a). What does this suggest about the upper bound relative to the corresponding probability?
(c) Suppose you will win \$$d$ if a fair coin flips heads and lose \$$d$ if it lands tails. Let X be the amount you get from a single coin flip. Compute $E(X)$ and $\text{SD}(X)$. What is the probability X will be less than one standard deviation from its mean value?
(d) Let X have three possible values, -1, 0, and 1, with probabilities $\frac{1}{18}$, $\frac{8}{9}$, and $\frac{1}{18}$ respectively. What is $P(|X - \mu| \geq 3\sigma)$, and how does it compare to the corresponding Chebyshev bound?
(e) Give a distribution for which $P(|X - \mu| \geq 5\sigma) = .04$.

48. For a discrete rv X taking values in $\{0, 1, 2, 3, \ldots\}$, we shall derive the following alternative formula for the mean:

$$\mu_X = \sum_{x=0}^{\infty} [1 - F(x)]$$

(a) Suppose for now the range of X is $\{0, 1, \ldots N\}$ for some positive integer N. By regrouping terms, show that

$$\sum_{x=0}^{N} [x \cdot p(x)] = p(1) + p(2) + p(3) + \cdots + p(N)$$
$$+ p(2) + p(3) + \cdots + p(N)$$
$$+ p(3) + \cdots + p(N)$$
$$\vdots$$
$$+ p(N)$$

(b) Rewrite each row in the above expression in terms of the cdf of X, and use this to establish that

$$\sum_{x=0}^{N} [x \cdot p(x)] = \sum_{x=0}^{N-1} [1 - F(x)]$$

(c) Let $N \to \infty$ in part (b) to establish the desired result, and explain why the resulting formula works even if the maximum value of X is finite. [*Hint:* If the largest possible value of X is N, what does $1 - F(x)$ equal for $x \geq N$?] (This derivation also implies that a discrete rv X has a finite mean iff the series $\sum [1 - F(x)]$ converges.)

(d) Let X have the pmf from Examples 2.10 and 2.19. Use the cdf of X and the alternative mean formula just derived to determine μ_X.

2.4   The Binomial Distribution

Many experiments conform either exactly or approximately to the following list of requirements:

1. The experiment consists of a sequence of n smaller experiments called *trials*, where n is fixed in advance of the experiment.
2. Each trial can result in one of the same two possible outcomes (dichotomous trials), which we denote by success (S) or failure (F).
3. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.
4. The probability of success is constant from trial to trial (homogeneous trials); we denote this probability by p.

DEFINITION

An experiment for which Conditions 1–4 are satisfied—a fixed number of dichotomous, independent, homogeneous trials—is called **a binomial experiment**.

Example 2.27 The same coin is tossed successively and independently *n* times. We arbitrarily use *S* to denote the outcome H (heads) and *F* to denote the outcome T (tails). Then this experiment satisfies Conditions 1–4. Tossing a thumbtack *n* times, with *S* = point up and *F* = point down, also results in a binomial experiment. ∎

Some experiments involve a sequence of independent trials for which there are more than two possible outcomes on any one trial. A binomial experiment can then be created by dividing the possible outcomes into two groups.

Example 2.28 The color of pea seeds is determined by a single genetic locus. If the two alleles at this locus are AA or Aa (the genotype), then the pea will be yellow (the phenotype), and if the allele is aa, the pea will be green. Suppose we pair off 20 Aa seeds and cross the two seeds in each of the ten pairs to obtain ten new genotypes. Call each new genotype a success *S* if it is aa and a failure otherwise. Then with this identification of *S* and *F*, the experiment becomes binomial with $n = 10$ and $p = P(\text{aa genotype})$. If each member of the pair is equally likely to contribute a or A, then $p = P(\text{a}) \cdot P(\text{a}) = (1/2)(1/2) = .25$. ∎

Example 2.29 A student has an iPod playlist containing 50 songs, of which 35 were recorded prior to the year 2015 and the other 15 were recorded more recently. Suppose the random play function is used to select five from among these 50 songs, without replacement, for listening during a walk between classes. Each selection of a song constitutes a trial; we regard a trial as a success if the selected song was recorded before 2015. Then clearly

$$P(S \text{ on first trial}) = \frac{35}{50} = .70$$

It may surprise you that the (unconditional) chance the second song is a success also equals .70! To see why, apply the Law of Total Probability:

$$\begin{aligned}
P(S \text{ on second trial}) &= P(SS \cup FS) \\
&= P(S \text{ on first})P(S \text{ on second}|S \text{ on first}) \\
&\quad + P(F \text{ on first})P(S \text{ on second}|F \text{ on first}) \\
&= \frac{35}{50} \cdot \frac{34}{49} + \frac{15}{50} \cdot \frac{35}{49} = \frac{35}{50}\left(\frac{34}{49} + \frac{15}{49}\right) = \frac{35}{50} = .70
\end{aligned}$$

Similarly, it can be shown that $P(S \text{ on } i\text{th trial}) = .70$ for $i = 3, 4, 5$, so the trials are homogeneous (Condition 4), with $p = .70$. However the trials are *not* independent (Condition 3), because for example,

$$P\left(S \text{ on fifth trial}\big|SSSS\right) = \frac{31}{46} = .67 \text{ whereas } P\left(S \text{ on fifth trial}\big|FFFF\right) = \frac{35}{46} = .76$$

(This matches our intuitive sense that later song selections "depend on" what was chosen before them.) The experiment is *not* binomial because the trials are not independent. In general, if sampling is without replacement, the experiment will not yield independent trials. If songs had been selected *with* replacement, then trials would have been independent, but this might have resulted in the same song being listened to more than once. ∎

Example 2.30 Suppose a state has 500,000 licensed drivers, of whom 400,000 are insured. A sample of 10 drivers is chosen without replacement. The *i*th trial is labeled *S* if the *i*th driver chosen is

insured. Although this situation would seem identical to that of Example 2.29, the important difference is that the size of the population being sampled is very large relative to the sample size. In this case

$$P\left(S \text{ on second}\middle|S \text{ on first}\right) = \frac{399,999}{499,999} \approx .80000$$

and

$$P\left(S \text{ on tenth}\middle|S \text{ on first nine}\right) = \frac{399,991}{499,991} = .799996 \approx .80000$$

These calculations suggest that although the trials are not exactly independent, the conditional probabilities differ so slightly from one another that for practical purposes the trials can be regarded as independent with constant $P(S) = .8$. Thus, to a very good approximation, the experiment is binomial with $n = 10$ and $p = .8$.   ∎

We will use the following convention in deciding whether a "without-replacement" experiment can be treated as being (approximately) binomial.

> **RULE**
> Consider sampling without replacement from a dichotomous population of size N. If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

By "analyzed," we mean that probabilities based on the binomial experiment assumptions will be quite close to the actual "without-replacement" probabilities, which are typically more difficult to calculate. In Example 2.29, $n/N = 5/50 = .1 > .05$, so the binomial experiment is not a good approximation, but in Example 2.30, $n/N = 10/500,000 < .05$.

2.4.1   The Binomial Random Variable and Distribution

In most binomial experiments, it is the total number of successes, rather than knowledge of exactly which trials yielded successes, that is of interest.

> **DEFINITION**
> Given a binomial experiment consisting of n trials, the **binomial random variable X** associated with this experiment is defined as
>
> $$X = \text{the number of successes among the } n \text{ trials}$$

Suppose, for example, that $n = 3$. Then there are eight possible outcomes for the experiment:

$$SSS \ SSF \ SFS \ SFF \ FSS \ FSF \ FFS \ FFF$$

From the definition of X, $X(SSF) = 2$, $X(SFF) = 1$, and so on. Possible values for X in an n-trial experiment are $x = 0, 1, 2, \ldots, n$.

NOTATION
We will write $X \sim \text{Bin}(n, p)$ to indicate that X is a binomial rv based on n trials with success probability p. Because the pmf of a binomial rv X depends on the two parameters n and p, we denote the pmf by $b(x; n, p)$.

Our next goal is to derive a formula for the binomial pmf. Consider first the case $n = 4$ for which each outcome, its probability, and corresponding x value are listed in Table 2.1. For example,

$$
\begin{aligned}
P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) && \text{independent trials} \\
&= p \cdot p \cdot (1 - p) \cdot p && \text{constant } P(S) \\
&= p^3 \cdot (1 - p)
\end{aligned}
$$

Table 2.1 Outcomes and probabilities for a binomial experiment with four trials

Outcome	x	Probability	Outcome	x	Probability
SSSS	4	p^4	FSSS	3	$p^3(1-p)$
SSSF	3	$p^3(1-p)$	FSSF	2	$p^2(1-p)^2$
SSFS	3	$p^3(1-p)$	FSFS	2	$p^2(1-p)^2$
SSFF	2	$p^2(1-p)^2$	FSFF	1	$p(1-p)^3$
SFSS	3	$p^3(1-p)$	FFSS	2	$p^2(1-p)^2$
SFSF	2	$p^2(1-p)^2$	FFSF	1	$p(1-p)^3$
SFFS	2	$p^2(1-p)^2$	FFFS	1	$p(1-p)^3$
SFFF	1	$p(1-p)^3$	FFFF	0	$(1-p)^4$

In this special case, we wish to determine $b(x; 4, p)$ for $x = 0, 1, 2, 3,$ and 4. For $b(3; 4, p)$, we identify which of the 16 outcomes yield an x value of 3 and sum the probabilities associated with each such outcome:

$$
b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1 - p)
$$

There are four outcomes with $x = 3$ and each has probability $p^3(1 - p)$; the probability depends only on the number of S's, *not* the order of S's and F's. So

$$
b(3; 4, p) = \left\{ \begin{array}{l} \text{number of outcomes} \\ \text{with } X = 3 \end{array} \right\} \cdot \left\{ \begin{array}{l} \text{probability of any particular} \\ \text{outcome with } X = 3 \end{array} \right\}
$$

Similarly, $b(2; 4, p) = 6p^2(1 - p)^2$, which is also the product of the number of outcomes with $X = 2$ and the probability of any such outcome.

In general,

$$
b(x; n, p) = \left\{ \begin{array}{l} \text{number of sequences of} \\ \text{length } n \text{ consisting of } x \text{ } S\text{'s} \end{array} \right\} \cdot \left\{ \begin{array}{l} \text{probability of any} \\ \text{particular such sequence} \end{array} \right\}
$$

Since the ordering of S's and F's is not important, the second factor in the previous equation is $p^x(1 - p)^{n-x}$ (for example, the first x trials resulting in S and the last $n - x$ resulting in F). The first factor is the number of ways of choosing x of the n trials to be S's—that is, the number of combinations of size x that can be constructed from n distinct objects (trials here).

> **THEOREM**
>
> $$b(x;n,p) = \begin{cases} \dbinom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Example 2.31 Each of six randomly selected cola drinkers is given a glass containing cola S and one containing cola F. The glasses are identical in appearance except for a code on the bottom to identify the cola. Suppose there is actually no tendency among cola drinkers to prefer one cola to the other. Then $p = P(\text{a selected individual prefers } S) = .5$, so with $X = $ the number among the six who prefer S, $X \sim \text{Bin}(6, .5)$.

Thus

$$P(X = 3) = b(3; 6, .5) = \binom{6}{3}(.5)^3(.5)^3 = 20(.5)^6 = .313$$

The probability that at least three prefer S is

$$P(X \geq 3) = \sum_{x=3}^{6} b(x; 6, .5) = \sum_{x=3}^{6} \binom{6}{x}(.5)^x(.5)^{6-x} = .656$$

and the probability that at most one prefers S is

$$P(X \leq 1) = \sum_{x=0}^{1} b(x; 6, .5) = .109 \qquad \blacksquare$$

2.4.2 Computing Binomial Probabilities

Even for a relatively small value of n, the computation of binomial probabilities can be tedious. Software and statistical tables are both available for this purpose; both are typically in terms of the cdf $F(x) = P(X \leq x)$ of the distribution, either in lieu of or in addition to the pmf. Various other probabilities can then be calculated using the proposition on cdfs from Sect. 2.2.

> **NOTATION**
> For $X \sim \text{Bin}(n, p)$, the cdf will be denoted by
>
> $$B(x; n, p) = P(X \leq x) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, \ldots, n$$

Table 2.2 at the end of this section provides the code for performing binomial calculations in both Matlab and R. In addition, Appendix Table A.1 tabulates the binomial cdf for $n = 5, 10, 15, 20, 25$ in combination with selected values of p.

Example 2.32 Suppose that 20% of all copies of a particular textbook fail a binding strength test. Let X denote the number among 15 randomly selected copies that fail the test. Then X has a binomial distribution with $n = 15$ and $p = .2$.

(a) The probability that at most 8 fail the test is

$$P(X \le 8) = \sum_{y=0}^{8} b(y; 15, .2) = B(8; 15, .2)$$

This is found at the intersection of the $p = .2$ column and $x = 8$ row in the $n = 15$ part of Table A.1: $B(8; 15, .2) = .999$. In Matlab, we may type `binocdf(8,15,.2)`; in R, the command is `pbinom(8,15,.2)`.

(b) The probability that exactly 8 fail is $P(X = 8) = b(8; 15, .2) = \binom{15}{8}(.2)^8(.8)^7 = .0034$. We can calculate this in Matlab or R with `binopdf(8,15,.2)` and `dbinom(8,15,.2)`, respectively. To use Table A.1, write

$$P(X = 8) = P(X \le 8) - P(X \le 7) = B(8; 15, .2) - B(7; 15, .2)$$

which is the difference between two consecutive entries in the $p = .2$ column. The result is $.999 - .996 = .003$.

(c) The probability that at least 8 fail is $P(X \ge 8) = 1 - P(X \le 7) = 1 - B(7; 15, .2)$. The cdf may be evaluated using Matlab or R as above, or by looking up the entry in the $x = 7$ row of the $p = .2$ column in Table A.1. In any case, we find $P(X \ge 8) = 1 - .996 = .004$.

(d) Finally, the probability that between 4 and 7, inclusive, fail is

$$\begin{aligned} P(4 \le X \le 7) &= P(X = 4, 5, 6, \text{ or } 7) = P(X \le 7) - P(X \le 3) \\ &= B(7; 15, .2) - B(3; 15, .2) = .996 - .648 = .348 \end{aligned}$$

Notice that this latter probability is the difference between the cdf values at $x = 7$ and $x = 3$, *not* $x = 7$ and $x = 4$. ∎

Example 2.33 An electronics manufacturer claims that at most 10% of its power supply units need service during the warranty period. To investigate this claim, technicians at a testing laboratory purchase 20 units and subject each one to accelerated testing to simulate use during the warranty period. Let p denote the probability that a power supply unit needs repair during the period (i.e., the proportion of *all* such units that need repair). The laboratory technicians must decide whether the data resulting from the experiment supports the claim that $p \le .10$. Let X denote the number among the 20 sampled that need repair, so $X \sim \text{Bin}(20, p)$. Consider the decision rule

Reject the claim that $p \le .10$ in favor of the conclusion that $p > .10$ if $x \ge 5$ (where x is the observed value of X), and consider the claim plausible if $x \le 4$

The probability that the claim is rejected when $p = .10$ (an incorrect conclusion) is

$$P(X \ge 5 \text{ when } p = .10) = 1 - B(4; 20, .1) = 1 - .957 = .043$$

The probability that the claim is not rejected when $p = .20$ (a different type of incorrect conclusion) is

$$P(X \le 4 \text{ when } p = .2) = B(4; 20, .2) = .630$$

The first probability is rather small, but the second is intolerably large. When $p = .20$, so that the manufacturer has grossly understated the percentage of units that need service, and the stated decision

rule is used, 63% of all samples of size 20 will result in the manufacturer's claim being judged plausible!

One might recognize that the probability of this second type of erroneous conclusion could be made smaller by changing the cutoff value 5 in the decision rule to something else. However, although replacing 5 by a smaller number would indeed yield a probability smaller than .630, the other probability would then increase. The only way to make both "error probabilities" small is to base the decision rule on an experiment involving many more units (i.e., to increase n). ∎

2.4.3   The Mean and Variance of a Binomial Random Variable

For $n = 1$, the binomial distribution becomes the Bernoulli distribution. From Example 2.17, the mean value of a Bernoulli variable is $\mu = p$, so the expected number of S's on any single trial is p. Since a binomial experiment consists of n trials, intuition suggests that for $X \sim \text{Bin}(n, p)$, $E(X) = np$, the product of the number of trials and the probability of success on a single trial. The expression for $\text{Var}(X)$ is not so obvious.

> **PROPOSITION**
>
> If $X \sim \text{Bin}(n, p)$, then $E(X) = np$, $\text{Var}(X) = np(1 - p) = npq$, and $\text{SD}(X) = \sqrt{npq}$ (where $q = 1 - p$).

Thus, calculating the mean and variance of a binomial rv does not necessitate evaluating summations of the sort we employed in Sect. 2.3. The proof of the result for $E(X)$ is sketched in Exercise 74.

Example 2.34 If 75% of all purchases at a store are made with a credit card and X is the number among ten randomly selected purchases made with a credit card, then $X \sim \text{Bin}(10, .75)$. Thus $E(X) = np = (10)(.75) = 7.5$, $\text{Var}(X) = np(1 - p) = 10(.75)(.25) = 1.875$, and $\sigma = \sqrt{1.875} = 1.37$. Again, even though X can take on only integer values, $E(X)$ need not be an integer. If we perform a large number of independent binomial experiments, each with $n = 10$ trials and $p = .75$, then the average number of S's per experiment will be close to 7.5. ∎

An important application of the binomial distribution is to estimating the precision of simulated probabilities, as in Sect. 1.6. The relative frequency definition of probability justified defining an estimate of a probability $P(A)$ by $\hat{P}(A) = X/n$, where n is the number of runs of the simulation program and X equals the number of runs in which event A occurred. Assuming the runs of our simulation are independent (and they usually are), the rv X has a binomial distribution with parameters n and $p = P(A)$. From the preceding proposition and the rescaling properties of mean and standard deviation, we have

$$E\big(\hat{P}(A)\big) = E\left(\frac{1}{n}X\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n}(np) = p = P(A)$$

Thus we expect the value of our estimate to coincide with the probability being estimated, in the sense that there is no reason for $\hat{P}(A)$ to be systematically higher or lower than $P(A)$. Also,

$$\mathrm{SD}\big(\hat{P}(A)\big) = \mathrm{SD}\left(\frac{1}{n}X\right) = \left|\frac{1}{n}\right| \cdot \mathrm{SD}(X) = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{P(A)[1-P(A)]}{n}} \quad (2.14)$$

Expression (2.14) is called the **standard error** of $\hat{P}(A)$ (essentially a synonym for standard deviation) and indicates the amount by which an estimate $\hat{P}(A)$ "typically" varies from the true probability $P(A)$. However, this expression isn't of much use in practice: we most often simulate a probability when $P(A)$ is unknown, which prevents us from using Eq. (2.14). As a solution, we simply substitute the estimate $\hat{P} = \hat{P}(A)$ into this expression and get

$$\mathrm{SD}\big(\hat{P}(A)\big) \approx \sqrt{\frac{\hat{P}\big(1-\hat{P}\big)}{n}}$$

This is the estimated standard error formula (1.8) given in Sect. 1.6. Very importantly, this estimated standard error gets closer to 0 as the number of runs, n, in the simulation increases.

2.4.4 Binomial Calculations with Software

Many software packages, including Matlab and R, have built-in functions to evaluate both the pmf and cdf of the binomial distribution (and many other named distributions). Table 2.2 summarizes the relevant code in both packages. The use of these functions was illustrated in Example 2.32.

Table 2.2 Binomial probability calculations in Matlab and R

Function:	pmf	cdf
Notation:	$b(x; n, p)$	$B(x; n, p)$
Matlab:	`binopdf(x, n, p)`	`binocdf(x, n, p)`
R:	`dbinom(x, n, p)`	`pbinom(x, n, p)`

2.4.5 Exercises: Section 2.4 (49–74)

49. Determine whether each of the following rvs has a binomial distribution. If it does, identify the values of the parameters n and p (if possible).
    (a) $X =$ the number of ⊡s in 10 rolls of a fair die
    (b) $X =$ the number of multiple-choice questions a student gets right on a 40-question test, when each question has four choices and the student is completely guessing
    (c) $X =$ the same as (b), but half the questions have four choices and the other half have three
    (d) $X =$ the number of women in a random sample of 8 students, from a class comprising 20 women and 15 men
    (e) $X =$ the total weight of 15 randomly selected apples
    (f) $X =$ the number of apples, out of a random sample of 15, that weigh more than 150 g
50. Compute the following binomial probabilities directly from the formula for $b(x; n, p)$:
    (a) $b(3; 8, .6)$
    (b) $b(5; 8, .6)$
    (c) $P(3 \le X \le 5)$ when $n = 8$ and $p = .6$
    (d) $P(1 \le X)$ when $n = 12$ and $p = .1$

51. Use Appendix Table A.1 or software to obtain the following probabilities:
    (a) $B(4; 10, .3)$
    (b) $b(4; 10, .3)$
    (c) $b(6; 10, .7)$
    (d) $P(2 \leq X \leq 4)$ when $X \sim \text{Bin}(10, .3)$
    (e) $P(2 \leq X)$ when $X \sim \text{Bin}(10, .3)$
    (f) $P(X \leq 1)$ when $X \sim \text{Bin}(10, .7)$
    (g) $P(2 < X < 6)$ when $X \sim \text{Bin}(10, .3)$
52. When circuit boards used in the manufacture of DVD players are tested, the long-run percentage of defectives is 5%. Let $X =$ the number of defective boards in a random sample of size $n = 25$, so $X \sim \text{Bin}(25, .05)$.
    (a) Determine $P(X \leq 2)$.
    (b) Determine $P(X \geq 5)$.
    (c) Determine $P(1 \leq X \leq 4)$.
    (d) What is the probability that none of the 25 boards is defective?
    (e) Calculate the expected value and standard deviation of X.
53. A company that produces fine crystal knows from experience that 10% of its goblets have cosmetic flaws and must be classified as "seconds."
    (a) Among six randomly selected goblets, how likely is it that only one is a second?
    (b) Among six randomly selected goblets, what is the probability that at least two are seconds?
    (c) If goblets are examined one by one, what is the probability that at most five must be selected to find four that are not seconds?
54. Suppose that only 25% of all drivers come to a complete stop at an intersection having flashing red lights in all directions when no other cars are visible. What is the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions,
    (a) At most 6 will come to a complete stop?
    (b) Exactly 6 will come to a complete stop?
    (c) At least 6 will come to a complete stop?
55. Refer to the previous exercise.
    (a) What is the expected number of drivers among the 20 that come to a complete stop?
    (b) What is the standard deviation of the number of drivers among the 20 that come to a complete stop?
    (c) What is the probability that the number of drivers among these 20 that come to a complete stop differs from the expected number by more than 2 standard deviations?
56. Suppose that 30% of all students who have to buy a text for a particular course want a new copy (the successes!), whereas the other 70% want a used copy. Consider randomly selecting 25 purchasers.
    (a) What are the mean value and standard deviation of the number who want a new copy of the book?
    (b) What is the probability that the number who want new copies is more than two standard deviations away from the mean value?
    (c) The bookstore has 15 new copies and 15 used copies in stock. If 25 people come in one by one to purchase this text, what is the probability that all 25 will get the type of book they want from current stock? [*Hint:* Let $X =$ the number who want a new copy. For what values of X will all 25 get what they want?]
    (d) Suppose that new copies cost $100 and used copies cost $70. Assume the bookstore has 50 new copies and 50 used copies. What is the expected value of total revenue from the sale of the next 25 copies purchased? [*Hint:* Let $h(X) =$ the revenue when X of the 25 purchasers want new copies. Express this as a linear function.]

57. Exercise 30 (Sect. 2.3) gave the pmf of Y, the number of traffic citations for a randomly selected individual insured by a company. What is the probability that among 15 randomly chosen such individuals
    (a) At least 10 have no citations?
    (b) Fewer than half have at least one citation?
    (c) The number that have at least one citation is between 5 and 10, inclusive?

58. A particular type of tennis racket comes in a midsize version and an oversize version. Sixty percent of all customers at a store want the oversize version.
    (a) Among ten randomly selected customers who want this type of racket, what is the probability that at least six want the oversize version?
    (b) Among ten randomly selected customers, what is the probability that the number who want the oversize version is within 1 standard deviation of the mean value?
    (c) The store currently has seven rackets of each version. What is the probability that all of the next ten customers who want this racket can get the version they want from current stock?

59. Twenty percent of all telephones of a certain type are submitted for service while under warranty. Of these, 60% can be repaired, whereas the other 40% must be replaced with new units. If a company purchases ten of these telephones, what is the probability that exactly two will end up being replaced under warranty?

60. The College Board reports that 2% of the two million high school students who take the SAT each year receive special accommodations because of documented disabilities (*Los Angeles Times*, July 16, 2002). Consider a random sample of 25 students who have recently taken the test.
    (a) What is the probability that exactly 1 received a special accommodation?
    (b) What is the probability that at least 1 received a special accommodation?
    (c) What is the probability that at least 2 received a special accommodation?
    (d) What is the probability that the number among the 25 who received a special accommodation is within 2 standard deviations of the number you would expect to be accommodated?
    (e) Suppose that a student who does not receive a special accommodation is allowed 3 hours for the exam, whereas an accommodated student is allowed 4.5 hours. What would you expect the average time allowed the 25 selected students to be?

61. Suppose that 90% of all batteries from a supplier have acceptable voltages. A certain type of flashlight requires two type-D batteries, and the flashlight will work only if both its batteries have acceptable voltages. Among ten randomly selected flashlights, what is the probability that at least nine will work? What assumptions did you make in the course of answering the question posed?

62. A *k-out-of-n system* functions provided that at least k of the n components function. Consider independently operating components, each of which functions (for the needed duration) with probability .96.
    (a) In a 3-component system, what is the probability that exactly two components function?
    (b) What is the probability a 2-out-of-3 system works?
    (c) What is the probability a 3-out-of-5 system works?
    (d) What is the probability a 4-out-of-5 system works?
    (e) What does the component probability (previously .96) need to equal so that the 4-out-of-5 system will function with probability at least .9999?

63. Bit transmission errors between computers sometimes occur, where one computer sends a 0 but the other computer receives a 1 (or vice versa). Because of this, the computer sending a message repeats each bit three times, so a 0 is sent as 000 and a 1 as 111. The receiving computer "decodes" each triplet by majority rule: whichever number, 0 or 1, appears more often in a triplet is declared to be the intended bit. For example, both 000 and 100 are decoded as 0, while 101 and

011 are decoded as 1. Suppose that 6% of bits are switched (0 to 1, or 1 to 0) during transmission between two particular computers, and that these errors occur independently during transmission.
  (a) Find the probability that a triplet is decoded incorrectly by the receiving computer.
  (b) Using your answer to part (a), explain how using triplets reduces communication errors.
  (c) How does your answer to part (a) change if each bit is repeated five times (instead of three)?
  (d) Imagine a 25 kilobit message (i.e., one requiring 25,000 bits to send). What is the expected number of errors if there is no bit repetition implemented? If each bit is repeated three times?

64. A very large batch of components has arrived at a distributor. The batch can be characterized as acceptable only if the proportion of defective components is at most .10. The distributor decides to randomly select 10 components and to accept the batch only if the number of defective components in the sample is at most 2.
  (a) What is the probability that the batch will be accepted when the actual proportion of defectives is .01? .05? .10? .20? .25?
  (b) Let p denote the actual proportion of defectives in the batch. A graph of P(batch is accepted) as a function of p, with p on the horizontal axis and P(batch is accepted) on the vertical axis, is called the *operating characteristic curve* for the acceptance sampling plan. Use the results of part (a) to sketch this curve for $0 \leq p \leq 1$.
  (c) Repeat parts (a) and (b) with "1" replacing "2" in the acceptance sampling plan.
  (d) Repeat parts (a) and (b) with "15" replacing "10" in the acceptance sampling plan.
  (e) Which of the three sampling plans, that of part (a), (c), or (d), appears most satisfactory, and why?

65. An ordinance requiring that a smoke detector be installed in all previously constructed houses has been in effect in a city for 1 year. The fire department is concerned that many houses remain without detectors. Let $p =$ the true proportion of such houses having detectors, and suppose that a random sample of 25 homes is inspected. If the sample strongly indicates that fewer than 80% of all houses have a detector, the fire department will campaign for a mandatory inspection program. Because of the costliness of the program, the department prefers not to call for such inspections unless sample evidence strongly argues for their necessity. Let X denote the number of homes with detectors among the 25 sampled. Consider rejecting the claim that $p \geq .8$ if $X \leq 15$.
  (a) What is the probability that the claim is rejected when the actual value of p is .8?
  (b) What is the probability of not rejecting the claim when $p = .7$? When $p = .6$?
  (c) How do the "error probabilities" of parts (a) and (b) change if the value 15 in the decision rule is replaced by 14?

66. A toll bridge charges $1.00 for passenger cars and $2.50 for other vehicles. Suppose that during daytime hours, 60% of all vehicles are passenger cars. If 25 vehicles cross the bridge during a particular daytime period, what is the resulting expected toll revenue? [*Hint*: Let $X =$ the number of passenger cars; then the toll revenue $h(X)$ is a linear function of X.]

67. A student who is trying to write a paper for a course has a choice of two topics, A and B. If topic A is chosen, the student will order two books through interlibrary loan, whereas if topic B is chosen, the student will order four books. The student believes that a good paper necessitates receiving and using at least half the books ordered for either topic chosen. If the probability that a book ordered through interlibrary loan actually arrives in time is .9 and books arrive independently of one another, which topic should the student choose to maximize the probability of writing a good paper? What if the arrival probability is only .5 instead of .9?

68. Twelve jurors are randomly selected from a large population. Each juror arrives at her or his conclusion about the case before the jury independently of the other jurors.

(a) In a criminal case, all 12 jurors must agree on a verdict. Let p denote the probability that a randomly selected member of the population would reach a guilty verdict based on the evidence presented (so a proportion $1 - p$ would reach "not guilty"). What is the probability, in terms of p, that the jury reaches a unanimous verdict one way or the other?

(b) For what values of p is the probability in part (a) the highest? For what value of p is the probability in (a) the lowest? Explain why this makes sense.

(c) In most civil cases, only a nine-person majority is required to decide a verdict. That is, if nine or more jurors favor the plaintiff, then the plaintiff wins; if at least nine jurors side with the defendant, then the defendant wins. Let p denote the probability that someone would side with the plaintiff based on the evidence. What is the probability, in terms of p, that the jury reaches a verdict one way or the other? How does this compare with your answer to part (a)?

69. Customers at a gas station pay with a credit card (A), debit card (B), or cash (C). Assume that successive customers make independent choices, with $P(A) = .5$, $P(B) = .2$, and $P(C) = .3$.

(a) Among the next 100 customers, what are the mean and variance of the number who pay with a debit card? Explain your reasoning.

(b) Answer part (a) for the number among the 100 who don't pay with cash.

70. An airport limousine can accommodate up to four passengers on any one trip. The company will accept a maximum of six reservations for a trip, and a passenger must have a reservation. From previous records, 20% of all those making reservations do not appear for the trip. In the following questions, assume independence, but explain why there could be dependence.

(a) If six reservations are made, what is the probability that at least one individual with a reservation cannot be accommodated on the trip?

(b) If six reservations are made, what is the expected number of available places when the limousine departs?

(c) Suppose the probability distribution of the number of reservations made is given in the accompanying table.

Number of reservations	3	4	5	6
Probability	.1	.2	.3	.4

Let X denote the number of passengers on a randomly selected trip. Obtain the probability mass function of X.

71. Let X be a binomial random variable with fixed n.

(a) Are there values of p ($0 \leq p \leq 1$) for which $\text{Var}(X) = 0$? Explain why this is so.

(b) For what value of p is $\text{Var}(X)$ maximized? [*Hint*: Either graph $\text{Var}(X)$ as a function of p or else take a derivative.]

72. (a) Show that $b(x; n, 1 - p) = b(n - x; n, p)$.

(b) Show that $B(x; n, 1 - p) = 1 - B(n - x - 1; n, p)$. [*Hint*: At most x S's is equivalent to at least $(n - x)$ F's.]

(c) What do parts (a) and (b) imply about the necessity of including values of p greater than .5 in Table A.1?

73. Refer to Chebyshev's inequality given in Sect. 2.3. Calculate $P(|X - \mu| \geq k\sigma)$ for $k = 2$ and $k = 3$ when $X \sim \text{Bin}(20, .5)$, and compare to the corresponding upper bounds. Repeat this for $X \sim \text{Bin}(20, .75)$.

74. Show that $E(X) = np$ when X is a binomial random variable. [*Hint*: Express $E(X)$ as a sum with lower limit $x = 1$. Then factor out np, let $y = x - 1$ so that the sum is from $y = 0$ to $y = n - 1$, and show that the sum equals 1.]

2.5   The Poisson Distribution

The binomial distribution was derived by starting with an experiment consisting of trials and applying the laws of probability to various outcomes of the experiment. There is no simple experiment on which the Poisson distribution is based, although we will shortly describe how it can be obtained from the binomial distribution by certain limiting operations.

DEFINITION

A random variable X is said to have a **Poisson distribution** with parameter μ ($\mu > 0$) if the pmf of X is

$$p(x;\mu) = \frac{e^{-\mu}\mu^x}{x!} \qquad x = 0, 1, 2, \ldots$$

We shall see shortly that μ is in fact the expected value of X, so the notation here is consistent with our previous use of the symbol μ. Because μ must be positive, $p(x;\mu) > 0$ for all possible x values. The fact that $\sum_{x=0}^{\infty} p(x;\mu) = 1$ is a consequence of the Maclaurin infinite series expansion of e^{μ}, which appears in most calculus texts:

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \tag{2.15}$$

If the two extreme terms in Eq. (2.15) are multiplied by $e^{-\mu}$ and then $e^{-\mu}$ is placed inside the summation, the result is

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

which shows that $p(x;\mu)$ fulfills the second condition necessary for specifying a pmf.

Example 2.35 Let X denote the number of creatures of a particular type captured in a trap during a given time period. Suppose that X has a Poisson distribution with $\mu = 4.5$, so on average traps will contain 4.5 creatures. [The article "Dispersal Dynamics of the Bivalve *Gemma gemma* in a Patchy Environment" (*Ecol. Monogr.*, 1995: 1–20) suggests this model; the bivalve *Gemma gemma* is a small clam.] The probability that a trap contains exactly five creatures is

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = .1708$$

The probability that a trap has at most five creatures is

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4.5}(4.5)^x}{x!} = e^{-4.5}\left[1 + 4.5 + \frac{4.5^2}{2!} + \cdots + \frac{4.5^5}{5!}\right] = .7029 \qquad \blacksquare$$

2.5.1   The Poisson Distribution as a Limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

> **PROPOSITION**
>
> Suppose that in the binomial pmf $b(x; n, p)$ we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n, p) \to p(x; \mu)$.

Proof Begin with the binomial pmf:

$$b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x} = \frac{n!}{x!(n - x)!} p^x (1 - p)^{n-x}$$
$$= \frac{n \cdot (n - 1) \cdot \cdots \cdot (n - x + 1)}{x!} p^x (1 - p)^{n-x}$$

Now multiply both the numerator and denominator by n^x:

$$b(x; n, p) = \frac{n}{n} \frac{n - 1}{n} \cdots \frac{n - x + 1}{n} \cdot \frac{(np)^x}{x!} \cdot \frac{(1 - p)^n}{(1 - p)^x}$$

Taking the limit as $n \to \infty$ and $p \to 0$ with $np \to \mu$,

$$\lim_{n \to \infty} b(x; n, p) = 1 \cdot 1 \cdots 1 \cdot \frac{\mu^x}{x!} \cdot \left(\lim_{n \to \infty} \frac{(1 - np/n)^n}{1} \right)$$

The limit on the right can be obtained from the calculus theorem that says the limit of $(1 - a_n/n)^n$ is e^{-a} if $a_n \to a$. Because $np \to \mu$,

$$\lim_{n \to \infty} b(x; n, p) = \frac{\mu^x}{x!} \cdot \lim_{n \to \infty} \left(1 - \frac{np}{n} \right)^n = \frac{\mu^x e^{-\mu}}{x!} = p(x; \mu)$$ ∎

According to the proposition, *in any binomial experiment for which the number of trials n is large and the success probability p is small, $b(x; n, p) \approx p(x; \mu)$ where $\mu = np$.* It is interesting to note that Siméon Poisson discovered the distribution that bears his name by this approach in the 1830s.

Table 2.3 shows the Poisson distribution for $\mu = 3$ along with three binomial distributions with $np = 3$, and Fig. 2.8 (from R) plots the Poisson along with the first two binomial distributions. The approximation is of limited use for $n = 30$, but of course the accuracy is better for $n = 100$ and much better for $n = 300$.

Table 2.3 Comparing the Poisson and three binomial distributions

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391	0.047553	0.049041	0.049787
1	0.141304	0.147070	0.148609	0.149361
2	0.227656	0.225153	0.224414	0.224042
3	0.236088	0.227474	0.225170	0.224042
4	0.177066	0.170606	0.168877	0.168031
5	0.102305	0.101308	0.100985	0.100819
6	0.047363	0.049610	0.050153	0.050409
7	0.018043	0.020604	0.021277	0.021604
8	0.005764	0.007408	0.007871	0.008102
9	0.001565	0.002342	0.002580	0.002701
10	0.000365	0.000659	0.000758	0.000810

Fig. 2.8 Comparing a
Poisson and two binomial
distributions

Legend:
- o Bin(30, .1)
- x Bin(100,.03)
- | Poisson(3)
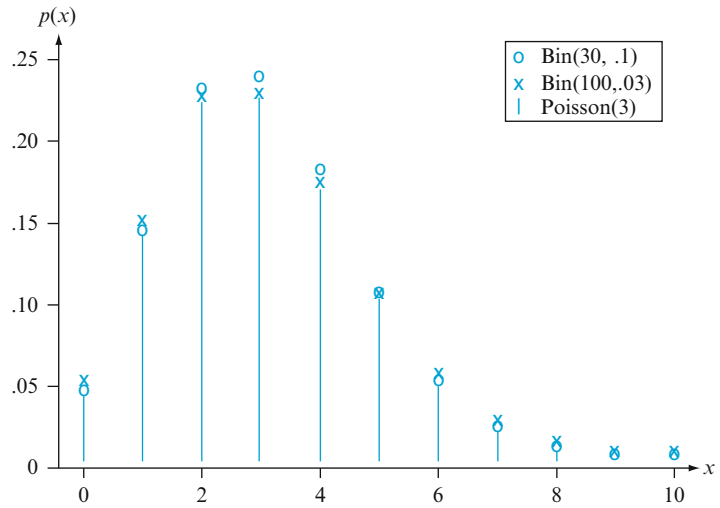
Example 2.36 Suppose you have a 4-megabit modem (4,000,000 bits/s) with bit error probability 10^{-8}. Assume bit errors occur independently, and assume your bit rate stays constant at 4 Mbps. What is the probability of exactly 3 bit errors in the next minute? Of at most 3 bit errors in the next minute?

Define a random variable X = the number of bit errors in the next minute. From the description, X satisfies the conditions of a binomial distribution; specifically, since a constant bit rate of 4 Mbps equates to 240,000,000 bits transmitted per minute, $X \sim \mathrm{Bin}(240000000, 10^{-8})$. Hence, the probability of exactly three bit errors in the next minute is

$$P(X = 3) = b\big(3; 240000000, 10^{-8}\big) = \binom{240000000}{3} \big(10^{-8}\big)^3 \big(1 - 10^{-8}\big)^{239999997}$$

For a variety of reasons, some calculators will struggle with this computation. The expression for the chance of at most 3 bit errors, $P(X \leq 3)$, is even worse. (The inability to compute such expressions in the nineteenth century, even with modest values of n and p, was Poisson's motive to derive an easily computed approximation.)

We may approximate these binomial probabilities using the Poisson distribution with $\mu = np = 240000000(10^{-8}) = 2.4$. Then

$$P(X = 3) \approx p(3; 2.4) = \frac{e^{-2.4}2.4^3}{3!} = .20901416$$

Similarly, the probability of at most 3 bit errors in the next minute is approximated by

$$P(X \leq 3) \approx \sum_{x=0}^{3} p(x; 2.4) = \sum_{x=0}^{3} \frac{e^{-2.4}2.4^x}{x!} = .77872291$$

Using modern software, the exact probabilities (i.e., using the binomial model) are .2090141655 and .7787229106, respectively. The Poisson approximations agree to eight decimal places and are clearly more computationally tractable. ∎

Many software packages will compute both $p(x; \mu)$ and the corresponding cdf $P(x; \mu)$ for specified values of x and μ upon request; the relevant Matlab and R functions appear in Table 2.4 at the end of

this section. Appendix Table A.2 exhibits the cdf $P(x; \mu)$ for $\mu = .1, .2, \ldots, 1, 2, \ldots, 10, 15,$ and 20. For example, if $\mu = 2$, then $P(X \leq 3) = P(3; 2) = .857$, whereas $P(X = 3) = P(3; 2) - P(2; 2) = .180$.

2.5.2   The Mean and Variance of a Poisson Random Variable

Since $b(x; n, p) \rightarrow p(x; \mu)$ as $n \rightarrow \infty, p \rightarrow 0, np \rightarrow \mu$, one might guess that the mean and variance of a binomial variable approach those of a Poisson variable. These limits are $np \rightarrow \mu$ and $np(1 - p) \rightarrow \mu$.

> **PROPOSITION**
> If X has a Poisson distribution with parameter μ, then $E(X) = \text{Var}(X) = \mu$.

These results can also be derived directly from the definitions of mean and variance (see Exercise 88 for the mean).

Example 2.37 (Example 2.35 continued) Both the expected number of creatures trapped and the variance of the number trapped equal 4.5, and $\sigma_X = \sqrt{\mu} = \sqrt{4.5} = 2.12$.   ∎

2.5.3   The Poisson Process

A very important application of the Poisson distribution arises in connection with the occurrence of events of a particular type over time. As an example, suppose that starting from a time point that we label $t = 0$, we are interested in counting the number of radioactive pulses recorded by a Geiger counter. If we make certain assumptions[2] about the way in which pulses occur—chiefly, that the number of pulses grows roughly linearly with time—then it can be shown that the number of pulses in any time interval of length t can be modeled by a Poisson distribution with mean $\mu = \lambda t$ for an appropriate positive constant λ. Since the expected number of pulses in an interval of length t is λt, the expected number in an interval of length 1 is λ. Thus λ is the long run number of pulses per unit of time.

If we replace "pulse" by "event," then the number of events occurring during a fixed time interval of length t has a Poisson distribution with parameter λt. Any process that has this distribution is called a **Poisson process**, and λ is called the *rate of the process*. Other examples of situations giving rise to a Poisson process include monitoring the status of a computer system over time, with breakdowns constituting the events of interest; recording the number of accidents in an industrial facility over time; answering 911 calls at a particular location; and observing the number of cosmic-ray showers from an observatory.

Example 2.36 hints at why this might be reasonable: if we "digitize" time—that is, divide time into discrete pieces, such as transmitted bits—and look at the number of the resulting time pieces that include an event, a binomial model is often applicable. If the number of time pieces is very large and the success probability close to zero, which would occur if we divided a fixed time frame into ever-smaller pieces, then we may invoke the Poisson approximation from earlier in this section.

---

[2] In Sect. 7.5, we present the formal assumptions required in this situation and derive the Poisson distribution that results from these assumptions.

Example 2.38   Suppose pulses arrive at the Geiger counter at an average rate of 6 per minute, so that $\lambda = 6$. To find the probability that in a 30-s interval at least one pulse is received, note that the number of pulses in such an interval has a Poisson distribution with parameter $\lambda t = 6(.5) = 3$ (.5 min is used because λ is expressed as a rate per minute). Then with $X =$ the number of pulses received in the 30-s interval,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-3}3^0}{0!} = .950$$

In a 1-h interval ($t = 60$), the expected number of pulses is $\mu = \lambda t = 6(60) = 360$, with a standard deviation of $\sigma = \sqrt{\mu} = \sqrt{360} = 18.97$. According to this model, in a typical hour we will observe 360 ± 19 pulses arrive at the Geiger counter.   ∎

Instead of observing events over time, consider observing events of some type that occur in a two- or three-dimensional region. For example, we might select on a map a certain region R of a forest, go to that region, and count the number of trees. Each tree would represent an event occurring at a particular point in space. Under appropriate assumptions (see Sect. 7.5), it can be shown that the number of events occurring in a region R has a Poisson distribution with parameter $\lambda \cdot a(R)$, where $a(R)$ is the area of R. The quantity λ is the expected number of events per unit area or volume.

2.5.4   Poisson Calculations with Software

Table 2.4 gives the Matlab and R commands for calculating Poisson probabilities.

Table 2.4 Poisson probability calculations

Function:	pmf	cdf
Notation:	$p(x; \mu)$	$P(x; \mu)$
Matlab:	`poisspdf(x, μ)`	`poisscdf(x, μ)`
R:	`dpois(x, μ)`	`ppois(x, μ)`

2.5.5   Exercises: Section 2.5 (75–89)

75. Let X, the number of flaws on the surface of a randomly selected carpet of a particular type, have a Poisson distribution with parameter $\mu = 5$. Use software or Appendix Table A.2 to compute the following probabilities:
    (a) $P(X \leq 8)$
    (b) $P(X = 8)$
    (c) $P(9 \leq X)$
    (d) $P(5 \leq X \leq 8)$
    (e) $P(5 < X < 8)$
76. Let X be the number of material anomalies occurring in a particular region of an aircraft gas-turbine disk. The article "Methodology for Probabilistic Life Prediction of Multiple-Anomaly Materials" (*Amer. Inst. of Aeronautics and Astronautics J.*, 2006: 787–793) proposes a Poisson distribution for X. Suppose $\mu = 4$.
    (a) Compute both $P(X \leq 4)$ and $P(X < 4)$.
    (b) Compute $P(4 \leq X \leq 8)$.

   (c)   Compute $P(8 \leq X)$.
   (d)   What is the probability that the observed number of anomalies exceeds the expected number by no more than one standard deviation?
77. Suppose that the number of drivers who travel between a particular origin and destination during a designated time period has a Poisson distribution with parameter $\mu = 20$ (suggested in the article "Dynamic Ride Sharing: Theory and Practice," *J. of Transp. Engr.*, 1997: 308–312). What is the probability that the number of drivers will
   (a)   Be at most 10?
   (b)   Exceed 20?
   (c)   Be between 10 and 20, inclusive? Be strictly between 10 and 20?
   (d)   Be within 2 standard deviations of the mean value?
78. Consider writing onto a computer disk and then sending it through a certifier that counts the number of missing pulses. Suppose this number X has a Poisson distribution with parameter $\mu = .2$. (Suggested in "Average Sample Number for Semi-Curtailed Sampling Using the Poisson Distribution," *J. Qual. Tech.*, 1983: 126–129.)
   (a)   What is the probability that a disk has exactly one missing pulse?
   (b)   What is the probability that a disk has at least two missing pulses?
   (c)   If two disks are independently selected, what is the probability that neither contains a missing pulse?
79. An article in the *Los Angeles Times* (Dec. 3, 1993) reports that 1 in 200 people carry the defective gene that causes inherited colon cancer. In a sample of 1000 individuals, what is the approximate distribution of the number who carry this gene? Use this distribution to calculate the approximate probability that
   (a)   Between 5 and 8 (inclusive) carry the gene.
   (b)   At least 8 carry the gene.
80. Suppose that only .10% of all computers of a certain type experience CPU failure during the warranty period. Consider a sample of 10,000 computers.
   (a)   What are the expected value and standard deviation of the number of computers in the sample that have the defect?
   (b)   What is the (approximate) probability that more than 10 sampled computers have the defect?
   (c)   What is the (approximate) probability that no sampled computers have the defect?
81. If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors? At most three pages with errors?
82. In proof testing of circuit boards, the probability that any particular diode will fail is .01. Suppose a circuit board contains 200 diodes.
   (a)   How many diodes would you expect to fail, and what is the standard deviation of the number that are expected to fail?
   (b)   What is the (approximate) probability that at least four diodes will fail on a randomly selected board?
   (c)   If five boards are shipped to a particular customer, how likely is it that at least four of them will work properly? (A board works properly only if all its diodes work.)
83. The article "Expectation Analysis of the Probability of Failure for Water Supply Pipes" (*J. Pipeline Syst. Eng. Pract.* 2012.3:36–46) recommends using a Poisson process to model the number of failures in commercial water pipes. The article also gives estimates of the failure

rate λ, in units of failures per 100 miles of pipe per day, for four different types of pipe and for many different years.

(a) For PVC pipe in 2008, the authors estimate a failure rate of 0.0081 failures per 100 miles of pipe per day. Consider a 100-mile-long segment of such pipe. What is the expected number of failures in 1 year (365 days)? Based on this expectation, what is the probability of at least one failure along such a pipe in 1 year?

(b) For cast iron pipe in 2005, the authors' estimate is $\lambda = 0.0864$ failures per 100 miles per day. Suppose a town had 1500 miles of cast iron pipe underground in 2005. What is the probability of at least one failure somewhere along this pipe system on any given day?

84. Organisms are present in ballast water discharged from a ship according to a Poisson process with a concentration of 10 organisms/m^3 (the article "Counting at Low Concentrations: The Statistical Challenges of Verifying Ballast Water Discharge Standards" (*Ecological Applications*, 2013: 339–351) considers using the Poisson process for this purpose).

(a) What is the probability that one cubic meter of discharge contains at least 8 organisms?

(b) What is the probability that the number of organisms in 1.5 m^3 of discharge exceeds its mean value by more than one standard deviation?

(c) For what amount of discharge would the probability of containing at least one organism be .999?

85. Suppose small aircraft arrive at an airport according to a Poisson process with rate $\lambda = 8$ per hour, so that the number of arrivals during a time period of t hours is a Poisson rv with parameter $\mu = 8t$.

(a) What is the probability that exactly 6 small aircraft arrive during a 1-h period? At least 6? At least 10?

(b) What are the expected value and standard deviation of the number of small aircraft that arrive during a 90-min period?

(c) What is the probability that at least 20 small aircraft arrive during a 2.5-h period? That at most 10 arrive during this period?

86. The number of people arriving for treatment at an emergency room can be modeled by a Poisson process with a rate parameter of five per hour.

(a) What is the probability that exactly four arrivals occur during a particular hour?

(b) What is the probability that at least four people arrive during a particular hour?

(c) How many people do you expect to arrive during a 45-min period?

87. Suppose that trees are distributed in a forest according to a two-dimensional Poisson process with rate λ, the expected number of trees per acre, equal to 80.

(a) What is the probability that in a certain quarter-acre plot, there will be at most 16 trees?

(b) If the forest covers 85,000 acres, what is the expected number of trees in the forest?

(c) Suppose you select a point in the forest and construct a circle of radius.1 mile. Let $X =$ the number of trees within that circular region. What is the pmf of X? [*Hint*: 1 sq mile $=$ 640 acres.]

88. Let X have a Poisson distribution with parameter μ. Show that $E(X) = \mu$ directly from the definition of expected value. [*Hint*: The first term in the sum equals 0, and then x can be canceled. Now factor out μ and show that what is left sums to 1.]

89. In some applications the distribution of a discrete rv X resembles the Poisson distribution except that zero is not a possible value of X. For example, let $X =$ the number of tattoos that an individual wants removed when s/he arrives at a tattoo removal facility. Suppose the pmf of X is

$$p(x) = k\frac{e^{-\theta}\theta^x}{x!} \quad x = 1, 2, 3, \ldots$$

(a) Determine the value of k. [*Hint:* The sum of all probabilities in the Poisson pmf is 1, and this pmf must also sum to 1.]

(b) If the mean value of X is 2.313035, what is the probability that an individual wants at most 5 tattoos removed?

(c) Determine the standard deviation of X when the mean value is as given in (b).

[*Note:* The article "An Exploratory Investigation of Identity Negotiation and Tattoo Removal" (*Academy of Marketing Science Review*, vol. 12, #6, 2008) gave a sample of 22 observations on the number of tattoos people wanted removed; estimates of μ and σ calculated from the data were 2.318182 and 1.249242, respectively.]

2.6   Other Discrete Distributions

This section introduces discrete distributions that are closely related to the binomial distribution. Whereas the binomial distribution is the *approximate* probability model for sampling without replacement from a finite dichotomous (S-F) population, the hypergeometric distribution is the exact probability model for the number of S's in the sample. The binomial rv X is the number of S's when the number n of trials is fixed, whereas the negative binomial distribution arises from fixing the number of S's desired and letting the number of trials be random.

2.6.1   The Hypergeometric Distribution

The assumptions leading to the hypergeometric distribution are as follows:

1. The population or set to be sampled consists of N individuals, objects, or elements (a *finite* population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

The random variable of interest is $X =$ the number of S's in the sample. The probability distribution of X depends on the parameters n, M, and N, so we wish to obtain the pmf $P(X = x) = h(x; n, M, N)$.

Example 2.39 During a particular period a university's information technology office received 20 service orders for problems with laptops, of which 8 were Macs and 12 were PCs. A sample of five of these service orders is to be selected for inclusion in a customer satisfaction survey. Suppose that the five are selected in a completely random fashion, so that any particular subset of size 5 has the same chance of being selected as does any other subset (think of putting the numbers 1, 2, ..., 20 on 20 identical slips of paper, mixing up the slips, and choosing five of them). What then is the probability that exactly 2 of the selected service orders were for PC laptops?

In this example, the population size is $N = 20$, the sample size is $n = 5$, and the number of S's (PC $= S$) and F's (Mac $= F$) in the population are $M = 12$ and $N - M = 8$, respectively. Let $X =$ the number of PCs among the five sampled service orders. Because all outcomes (each consisting of five particular orders) are equally likely,

$$P(X = 2) = h(2; 5, 12, 20) = \frac{\text{number of outcomes having } X = 2}{\text{number of possible outcomes}}$$

The number of possible outcomes in the experiment is the number of ways of selecting 5 from the 20 objects without regard to order—that is, $\binom{20}{5}$. To count the number of outcomes having $X = 2$, note that there are $\binom{12}{2}$ ways of selecting two of the PC orders, and for each such way there are $\binom{8}{3}$ ways of selecting the three Mac orders to fill out the sample. The Fundamental Counting Principle from Sect. 1.3 then gives $\binom{12}{2} \cdot \binom{8}{3}$ as the number of outcomes with $X = 2$, so

$$h(2; 5, 12, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = .238$$

∎

In general, if the sample size n is smaller than the number of successes in the population (M), then the largest possible X value is n. However, if $M < n$ (e.g., a sample size of 25 and only 15 successes in the population), then X can be at most M. Similarly, whenever the number of population failures ($N - M$) exceeds the sample size, the smallest possible X value is 0 (since all sampled individuals might then be failures). However, if $N - M < n$, the smallest possible X value is $n - (N - M)$. Summarizing, the possible values of X satisfy the restriction $\max(0, n - N + M) \le x \le \min(n, M)$. An argument parallel to that of the previous example gives the pmf of X.

> **PROPOSITION**
>
> If X is the number of S's in a random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the **hypergeometric distribution**, is given by
>
> $$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N - M}{n - x}}{\binom{N}{n}} \qquad (2.16)$$
>
> for x an integer satisfying $\max(0, n - N + M) \le x \le \min(n, M)$.[3]

---

[3] If we define $\binom{a}{b} = 0$ for $a < b$, then $h(x; n, M, N)$ may be applied for all integers $0 \le x \le n$.

In Example 2.39, $n = 5$, $M = 12$, and $N = 20$, so $h(x; 5, 12, 20)$ for $x = 0, 1, 2, 3, 4, 5$ can be obtained by substituting these numbers into Eq. (2.16).

Example 2.40 *Capture–recapture*. Five individuals from an animal population thought to be near extinction in a region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of 10 of these animals is selected. Let $X =$ the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that (a) $X = 2$? (b) $X \leq 2$?

Application of the hypergeometric distribution here requires assuming that every subset of ten animals has the same chance of being captured. This in turn implies that released animals are no easier or harder to catch than are those not initially captured. Then the parameter values are $n = 10$, $M = 5$ (five tagged animals in the population), and $N = 25$, so

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}} \qquad x = 0, 1, 2, 3, 4, 5$$

For part (a),

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

For part (b),

$$P(X \leq 2) = P(X = 0, 1, \text{ or } 2) = \sum_{x=0}^{2} h(x; 10, 5, 25)$$
$$= .057 + .257 + .385 = .699 \qquad\blacksquare$$

Matlab, R, and other software packages will easily generate hypergeometric probabilities; see Table 2.5 at the end of this section. Comprehensive tables of the hypergeometric distribution are available, but because the distribution has three parameters, these tables require much more space than tables for the binomial distribution.

As in the binomial case, there are simple expressions for $E(X)$ and $\text{Var}(X)$ for hypergeometric rvs.

PROPOSITION

The mean and variance of the hypergeometric rv X having pmf $h(x; n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N} \quad \text{Var}(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

The ratio M/N is the proportion of S's in the population. Replacing M/N by p in $E(X)$ and $\text{Var}(X)$ gives

$$E(X) = np \tag{2.17}$$

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

Expression (2.17) shows that the means of the binomial and hypergeometric rvs are equal, whereas the variances of the two rvs differ by the factor $(N-n)/(N-1)$, often called the **finite population correction factor**. This factor is less than 1, so the hypergeometric variable has smaller variance than does the binomial rv. The correction factor can be written $(1 - n/N)/(1 - 1/N)$, which is approximately 1 when n is small relative to N.

Example 2.41 (Example 2.40 continued) In the animal-tagging example, $n = 10$, $M = 5$, and $N = 25$, so $p = \frac{5}{25} = .2$ and

$$E(X) = 10(.2) = 2$$

$$\text{Var}(X) = \frac{25-10}{25-1}(10)(.2)(.8) = (.625)(1.6) = 1$$

If the sampling were carried out *with* replacement, $\text{Var}(X) = 1.6$.

Suppose the population size N is not actually known, so the value x is observed and we wish to estimate N. It is reasonable to equate the observed sample proportion of S's, x/n, with the population proportion, M/N, giving the estimate

$$\hat{N} = \frac{M \cdot n}{x}$$

For example, if $M = 100$, $n = 40$, and $x = 16$, then $\hat{N} = 250$.   ∎

Our rule in Sect. 2.4 stated that if sampling is without replacement but n/N is at most .05, then the binomial distribution can be used to compute approximate probabilities involving the number of S's in the sample. A more precise statement is as follows: Let the population size, N, and number of population S's, M, get large with the ratio M/N approaching p. Then $h(x; n, M, N)$ approaches the binomial pmf $b(x; n, p)$; so for n/N small, the two are approximately equal provided that p is not too near either 0 or 1. This is the rationale for our rule.

2.6.2   The Negative Binomial and Geometric Distributions

The negative binomial distribution is based on an experiment satisfying the following conditions:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so $P(S$ on trial $i) = p$ for $i = 1, 2, 3 \ldots$.
4. The experiment continues (trials are performed) until a total of r successes has been observed, where r is a specified positive integer.

The random variable of interest is $X =$ the number of trials required to achieve the rth success, and X is called a **negative binomial random variable**. In contrast to the binomial rv, the number of

successes is fixed and the number of *trials* is random. Possible values of X are $r, r + 1, r + 2, \ldots$, since it takes at least r trials to achieve r successes.

Let $nb(x; r, p)$ denote the pmf of X. The event $\{X = x\}$ is equivalent to $\{r - 1$ S's in the first $(x - 1)$ trials and an S on the xth trial$\}$, e.g., if $r = 5$ and $x = 15$, then there must be four S's in the first 14 trials and trial 15 must be an S. Since trials are independent,

$$nb(x; r, p) = P(X = x) = P(r - 1 \; S\text{'s on the first } x - 1 \text{ trials}) \cdot P(S) \qquad (2.18)$$

The first probability on the far right of Eq. (2.18) is the binomial probability

$$\binom{x - 1}{r - 1} p^{r-1}(1 - p)^{(x-1)-(r-1)} \quad \text{where } P(S) = p$$

Simplifying and then multiplying by the extra factor of p at the end of Eq. (2.18) yields the following.

PROPOSITION

The pmf of the negative binomial rv X with parameters $r =$ desired number of S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x - 1}{r - 1} p^r (1 - p)^{x-r} \quad x = r, r + 1, r + 2, \ldots$$

Example 2.42 A pediatrician wishes to recruit four couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p = P$(a randomly selected couple agrees to participate). If $p = .2$, what is the probability that exactly 15 couples must be asked before 4 are found who agree to participate? Substituting $r = 4$, $p = .2$, and $x = 15$ into $nb(x; r, p)$ gives

$$nb(15; 4, 2) = \binom{15 - 1}{4 - 1} .2^4 .8^{11} = .050$$

The probability that at most 15 couples need to be asked is

$$P(X \le 15) = \sum_{x=4}^{15} nb(x; 4, .2) = \sum_{x=4}^{15} \binom{x - 1}{3} .2^4 .8^{x-4} = .352 \qquad \blacksquare$$

In the special case $r = 1$, the pmf is

$$nb(x; 1, p) = (1 - p)^{x-1} p \quad x = 1, 2, \ldots \qquad (2.19)$$

In Example 2.10, we derived the pmf for the number of trials necessary to obtain the first S, and the pmf there is identical to Eq. (2.19). The random variable $X =$ number of trials required to achieve one success is referred to as a **geometric random variable**, and the pmf in Eq. (2.19) is called the **geometric distribution**. The name is appropriate because the probabilities constitute a geometric series: $p, (1 - p)p, (1 - p)^2 p, \ldots$. To see that the sum of the probabilities is 1, recall that the sum of a geometric series is $a + ar + ar^2 + \ldots = a/(1 - r)$ if $|r| < 1$, so for $p > 0$,

$$p + (1-p)p + (1-p)^2 p + \cdots = \frac{p}{1-(1-p)} = 1$$

In Example 2.19, the expected number of trials until the first S was shown to be $1/p$. Intuitively, we would then expect to need $r \cdot 1/p$ trials to achieve the rth S, and this is indeed $E(X)$. There is also a simple formula for $\text{Var}(X)$.

> **PROPOSITION**
>
> If X is a negative binomial rv with parameters r and p, then
>
> $$E(X) = \frac{r}{p} \quad \text{Var}(X) = \frac{r(1-p)}{p^2}$$

Example 2.43 (Example 2.42 continued) With $p = .2$, the expected number of couples the doctor must speak to in order to find 4 that will agree to participate is $r/p = 4/.2 = 20$. This makes sense, since with $p = .2 = 1/5$ it will take five attempts, on average, to achieve one success. The corresponding variance is $4(1 - .2)/(.2)^2 = 80$, for a standard deviation of about 8.9.   ∎

Since they are based on similar experiments, some caution must be taken to distinguish the binomial and negative binomial models, as seen in the next example.

Example 2.44 In many communication systems, a receiver will send a short signal back to the transmitter to indicate whether a message has been received correctly or with errors. (These signals are often called an *acknowledgement* and a *non-acknowledgement*, respectively. Bit sum checks and other tools are used by the receiver to determine the absence or presence of errors.) Assume we are using such a system in a noisy channel, so that each message is sent error-free with probability .86, independent of all other messages. What is the probability that in 10 transmissions, exactly 8 will succeed? What is the probability the system will require exactly 10 attempts to successfully transmit 8 messages?

While these two questions may sound similar, they require two different models for solution. To answer the first question, let X represent the number of successful transmissions out of 10. Then $X \sim \text{Bin}(10, .86)$, and the answer is

$$P(X = 8) = b(8; 10, .86) = \binom{10}{8}(.86)^8(.14)^2 = .2639$$

However, the event {exactly 10 attempts required to successfully transmit 8 messages} is more restrictive: not only must we observe 8 S's and 2 F's in 10 trials, but *the last trial must be a success*. Otherwise, it took fewer than 10 tries to send 8 messages successfully. Define a variable $Y = $ the number of transmissions (trials) required to successfully transmit 8 messages. Then Y is negative binomial, with $r = 8$ and $p = .86$, and the answer to the second question is

$$P(Y = 10) = nb(10; 8, .86) = \binom{10-1}{8-1}(.86)^8(.14)^2 = .2111$$

Notice this is smaller than the answer to the first question, which makes sense because (as we noted) the second question imposes an additional constraint. In fact, you can think of the "-1" terms in the negative binomial pmf as accounting for this loss of flexibility in the placement of S's and F's.

Similarly, the expected number of successful transmissions in 10 attempts is $E(X) = np = 10(.86)$ = 8.6, while the expected number of attempts required to successfully transmit 8 messages is $E(Y) = r/p = 8/.86 = 9.3$. In the first case, the number of trials ($n = 10$) is fixed, while in the second case the desired number of successes ($r = 8$) is fixed.   ∎

By expanding the binomial coefficient in front of $p^r(1 - p)^{x-r}$ and doing some cancellation, it can be seen that $nb(x; r, p)$ is well-defined even when r is not an integer. This *generalized negative binomial distribution* has been found to fit observed data quite well in a wide variety of applications.

2.6.3 Alternative Definition of the Negative Binomial Distribution

There is not universal agreement on the definition of a negative binomial random variable (or, by extension, a geometric rv). It is not uncommon in the literature, as well as in some textbooks, to see the number of *failures* preceding the rth success called "negative binomial"; in our notation, this simply equals $X - r$. Possible values of this "number of failures" variable are 0, 1, 2, …... Similarly, the geometric distribution is sometimes defined in terms of the number of failures preceding the first success in a sequence of independent and identical trials. If one uses these alternative definitions, then the pmf and mean formula must be adjusted accordingly. (The variance, however, will stay the same.)

The developers of Matlab and R are among those who have adopted this alternative definition; as a result, we must be careful with our inputs to the relevant software functions. The pmf syntax for the distributions in this section are cataloged in Table 2.5; cdfs may be invoked by changing `pdf` to `cdf` in Matlab or the initial letter `d` to `p` in R. Notice the input argument $x - r$ for the negative binomial functions: both software packages request the number of failures, rather than the number of trials.

Table 2.5 Matlab and R code for hypergeometric and negative binomial calculations

	Hypergeometric	Negative Binomial
Function:	pmf	pmf
Notation:	$h(x; n, M, N)$	$nb(x; r, p)$
Matlab:	`hygepdf(x, N, M, n)`	`nbinpdf(x − r, r, p)`
R:	`dhyper(x, M, N − M, n)`	`dnbinom(x − r, r, p)`

For example, suppose X has a hypergeometric distribution with $n = 10$, $M = 5$, $N = 25$ as in Example 2.40. Using Matlab, we may calculate $P(X = 2) =$ `hygepdf(2,25,5,10)` and $P(X \leq 2)$ = `hygecdf(2,25,5,10)`. The corresponding R function calls are `dhyper(2,5,20,10)` and `phyper(2,5,20,10)`, respectively. If X is the negative binomial variable of Example 2.42 with parameters $r = 4$ and $p = .2$, then the chance of requiring 15 trials to achieve 4 successes (i.e., 11 total failures) can be found in Matlab with `nbinpdf(11,4, .2)` and in R using the command `dnbinom(11,4, .2)`.

2.6.4 Exercises: Section 2.6 (90–106)

90. An electronics store has received a shipment of 20 table radios that have connections for an iPod or iPhone. Twelve of these have two slots (so they can accommodate both devices), and the other eight have a single slot. Suppose that six of the 20 radios are randomly selected to be stored under a shelf where radios are displayed, and the remaining ones are placed in a

storeroom. Let $X =$ the number among the radios stored under the display shelf that have two slots.
   (a)   What kind of a distribution does X have (name and values of all parameters)?
   (b)   Compute $P(X = 2)$, $P(X \leq 2)$, and $P(X \geq 2)$.
   (c)   Calculate the mean value and standard deviation of X.

91. Each of 12 refrigerators has been returned to a distributor because of an audible, high-pitched, oscillating noise when the refrigerator is running. Suppose that 7 of these refrigerators have a defective compressor and the other 5 have less serious problems. If the refrigerators are examined in random order, let X be the number among the first 6 examined that have a defective compressor. Compute the following:
   (a) $P(X = 5)$
   (b) $P(X \leq 4)$
   (c)   The probability that X exceeds its mean value by more than 1 standard deviation.
   (d)   Consider a large shipment of 400 refrigerators, of which 40 have defective compressors. If X is the number among 15 randomly selected refrigerators that have defective compressors, describe a less tedious way to calculate (at least approximately) $P(X \leq 5)$ than to use the hypergeometric pmf.

92. An instructor who taught two sections of statistics last term, the first with 20 students and the second with 30, decided to assign a term project. After all projects had been turned in, the instructor randomly ordered them before grading. Consider the first 15 graded projects.
   (a)   What is the probability that exactly 10 of these are from the second section?
   (b)   What is the probability that at least 10 of these are from the second section?
   (c)   What is the probability that at least 10 of these are from the same section?
   (d)   What are the mean and standard deviation of the number among these 15 that are from the second section?
   (e)   What are the mean and standard deviation of the number of projects not among these first 15 that are from the second section?

93. A geologist has collected 10 specimens of basaltic rock and 10 specimens of granite. The geologist instructs a laboratory assistant to randomly select 15 of the specimens for analysis.
   (a)   What is the pmf of the number of granite specimens selected for analysis?
   (b)   What is the probability that all specimens of one of the two types of rock are selected for analysis?
   (c)   What is the probability that the number of granite specimens selected for analysis is within 1 standard deviation of its mean value?

94. A personnel director interviewing 11 senior engineers for four job openings has scheduled six interviews for the first day and five for the second day of interviewing. Assume the candidates are interviewed in random order.
   (a)   What is the probability that x of the top four candidates are interviewed on the first day?
   (b)   How many of the top four candidates can be expected to be interviewed on the first day?

95. Twenty pairs of individuals playing in a bridge tournament have been seeded 1, ..., 20. In the first part of the tournament, the 20 are randomly divided into 10 east–west pairs and 10 north–south pairs.
   (a)   What is the probability that x of the top 10 pairs end up playing east–west?
   (b)   What is the probability that all of the top five pairs end up playing the same direction?
   (c)   If there are $2n$ pairs, what is the pmf of $X =$ the number among the top n pairs who end up playing east–west? What are $E(X)$ and $\text{Var}(X)$?

96. A second-stage smog alert has been called in an area of Los Angeles County in which there are 50 industrial firms. An inspector will visit 10 randomly selected firms to check for violations of regulations.
    (a) If 15 of the firms are actually violating at least one regulation, what is the pmf of the number of firms visited by the inspector that are in violation of at least one regulation?
    (b) If there are 500 firms in the area, of which 150 are in violation, approximate the pmf of part (a) by a simpler pmf.
    (c) For X = the number among the 10 visited that are in violation, compute $E(X)$ and Var(X) both for the exact pmf and the approximating pmf in part (b).

97. A shipment of 20 integrated circuits (ICs) arrives at an electronics manufacturing site. The site manager will randomly select 4 ICs and test them to see whether they are faulty. Unknown to the site manager, 5 of these 20 ICs are faulty.
    (a) Suppose the shipment will be accepted if and only if none of the inspected ICs is faulty. What is the probability this shipment of 20 ICs will be accepted?
    (b) Now suppose the shipment will be accepted if and only if at most one of the inspected ICs is faulty. What is the probability this shipment of 20 ICs will be accepted?
    (c) How do your answers to (a) and (b) change if the number of faculty ICs in the shipment is 3 instead of 5? Recalculate (a) and (b) to verify your claim.

98. Suppose that 20% of all individuals have an adverse reaction to a particular drug. A medical researcher will administer the drug to one individual after another until the first adverse reaction occurs. Define an appropriate random variable and use its distribution to answer the following questions.
    (a) What is the probability that when the experiment terminates, four individuals have not had adverse reactions?
    (b) What is the probability that the drug is administered to exactly five individuals?
    (c) What is the probability that at most four individuals do not have an adverse reaction?
    (d) How many individuals would you expect to not have an adverse reaction, and how many individuals would you expect to be given the drug?
    (e) What is the probability that the number of individuals given the drug is within one standard deviation of what you expect?

99. Suppose that $p = P$(female birth) = .5. A couple wishes to have exactly two female children in their family. They will have children until this condition is fulfilled.
    (a) What is the probability that the family has x male children?
    (b) What is the probability that the family has four children?
    (c) What is the probability that the family has at most four children?
    (d) How many children would you expect this family to have? How many male children would you expect this family to have?

100. A family decides to have children until it has three children of the same gender. Assuming $P(B)$ = $P(G)$ = .5, what is the pmf of X = the number of children in the family?

101. Three brothers and their wives decide to have children until each family has two female children. Let X = the total number of male children born to the brothers. What is $E(X)$, and how does it compare to the expected number of male children born to each brother?

102. According to the article "Characterizing the Severity and Risk of Drought in the Poudre River, Colorado" (*J. of Water Res. Planning and Mgmnt.*, 2005: 383–393), the drought length Y is the number of consecutive time intervals in which the water supply remains below a critical value y_0 (a deficit), preceded and followed by periods in which the supply exceeds this value (a surplus). The cited paper proposes a geometric distribution with p = .409 for this random variable.

    (a) What is the probability that a drought lasts exactly 3 intervals? At least 3 intervals?

    (b) What is the probability that the length of a drought exceeds its mean value by at least one standard deviation?

103. Individual A has a red die and B has a green die (both fair). If they each roll until they obtain five "doubles" ($\square\square$, ..., $\boxplus\boxplus$), what is the pmf of $X = $ the total number of times a die is rolled? What are $E(X)$ and SD(X)?

104. A carnival game consists of spinning a wheel with 10 slots, nine red and one blue. If you land on the blue slot, you win a prize. Suppose your significant other *really* wants that prize, so you will play until you win.

    (a) What is the probability you'll win on the first spin?

    (b) What is the probability you'll require exactly 5 spins? At least 5 spins? At most five spins?

    (c) What is the expected number of spins required for you to win the prize, and what is the corresponding standard deviation?

105. A kinesiology professor, requiring volunteers for her study, approaches students one by one at a campus hub. She will continue until she acquires 40 volunteers. Suppose that 25% of students are willing to volunteer for the study, that the professor's selections are random, and that the student population is large enough that individual "trials" (asking a student to participate) may be treated as independent.

    (a) What is the expected number of students the kinesiology professor will need to ask in order to get 40 volunteers? What is the standard deviation?

    (b) Determine the probability that the number of students the kinesiology professor will need to ask is within one standard deviation of the mean.

106. Refer back to the communication system of Example 2.44. Suppose a voice packet can be transmitted a maximum of 10 times, i.e., if the 10th attempt fails, no 11th attempt is made to retransmit the voice packet. Let $X = $ the number of times a message is transmitted. Assuming each transmission succeeds with probability p, determine the pmf of X. Then obtain an expression for the expected number of times a packet is transmitted.

2.7 Moments and Moment Generating Functions

The expected values of integer powers of X and $X - \mu$ are often referred to as *moments*, terminology borrowed from physics. In this section, we'll discuss the general topic of moments and develop a shortcut for computing them.

> **DEFINITION**
> The **kth moment** of a random variable X is $E(X^k)$, while the **kth moment about the mean** (or **kth central moment**) of X is $E[(X - \mu)^k]$, where $\mu = E(X)$.

For example, $\mu = E(X)$ is the "first moment" of X and corresponds to the center of mass of the distribution of X. Similarly, Var(X) $= E[(X - \mu)^2]$ is the second moment of X about the mean, which is known in physics as the *moment of inertia*.

Example 2.45 A popular brand of dog food is sold in 5, 10, 15, and 20 lb bags. Let X be the weight of the next bag purchased, and suppose the pmf of X is

x	5	10	15	20
$p(x)$	.1	.2	.3	.4

The first moment of X is its mean:

$$\mu = E(X) = \sum_{x \in D} xp(x) \quad = 5(.1) + 10(.2) + 15(.3) + 20(.4) = 15\,\text{lbs}$$

The second moment about the mean is the variance:

$$\sigma^2 = E\left[(X - \mu)^2\right] = \sum_{x \in D} (x - \mu)^2 p(x)$$

$$= (5 - 15)^2(.1) + (10 - 15)^2(.2) + (15 - 15)^2(.3) + (20 - 15)^2(.4) = 25,$$

for a standard deviation of 5 lb. The third central moment of X is

$$E\left[(X - \mu)^3\right] = \sum_{x \in D} (x - \mu)^3 p(x)$$

$$= (5 - 15)^3(.1) + (10 - 15)^3(.2) + (15 - 15)^3(.3) + (20 - 15)^3(.4) = -75$$

We'll discuss an interpretation of this last number next.   ∎

It is not difficult to verify that the third moment about the mean is 0 if the pmf of X is symmetric. So, we would like to use $E[(X - \mu)^3]$ as a measure of lack of symmetry, but it depends on the scale of measurement. If we switch the unit of weight in Example 2.45 from pounds to ounces or kilograms, the value of the third moment about the mean (as well as the values of all the other moments) will change. But we can achieve scale independence by dividing the third moment about the mean by σ^3:

$$\frac{E\left[(X - \mu)^3\right]}{\sigma^3} = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] \tag{2.20}$$

Expression (2.20) is our measure of departure from symmetry, called the **skewness coefficient**. The skewness coefficient for a symmetric distribution is 0 because its third moment about the mean is 0. However, in the foregoing example the skewness coefficient is $E[(X - \mu)^3]/\sigma^3 = -75/5^3 = -0.6$. When the skewness coefficient is negative, as it is here, we say that the distribution is *negatively skewed* or that it is *skewed to the left*. Generally speaking, it means that the distribution stretches farther to the left of the mean than to the right.

If the skewness were positive, then we would say that the distribution is *positively skewed* or that it is *skewed to the right*. For example, reverse the order of the probabilities in the $p(x)$ table above, so the probabilities of the values 5, 10, 15, 20 are now .4, .3, .2, and .1, (customers now favor much smaller bags of dog food). Exercise 119 shows that this changes the sign but not the magnitude of the skewness coefficient, so it becomes +0.6 and the distribution is skewed right. Both distributions are illustrated in Fig. 2.9.
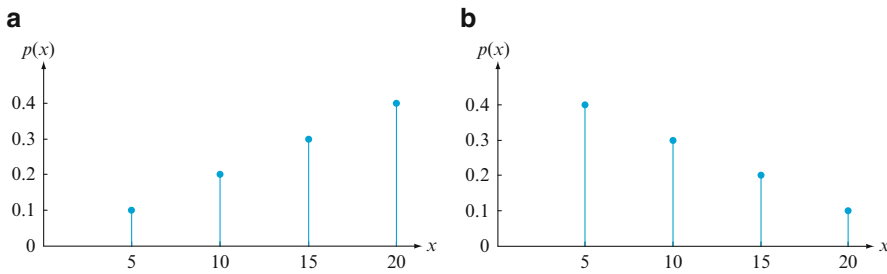
Fig. 2.9 Departures from symmetry: (**a**) skewness coefficient < 0 (skewed left); (**b**) skewness coefficient > 0 (skewed right)

2.7.1   The Moment Generating Function

Calculation of the mean, variance, skewness coefficient, etc. for a particular discrete rv requires extensive, sometimes tedious, summation. Mathematicians have developed a tool, the moment generating function, that will allow us to determine the moments of a distribution with less effort. Moreover, this function will allow us to derive properties of several of our major probability distributions here and in subsequent sections of the book.

> **DEFINITION**
> The **moment generating function** (**mgf**) of a discrete random variable X is defined to be
>
> $$M_X(t) = E\left(e^{tX}\right) = \sum_{x \in D} e^{tx} p(x)$$
>
> where D is the set of possible X values. The moment generating function exists iff $M_X(t)$ is defined for an interval that includes zero as well as positive and negative values of t.

For any random variable X, the mgf evaluated at $t = 0$ is

$$M_X(0) = E\left(e^{0X}\right) = \sum_{x \in D} e^{0x} p(x) = \sum_{x \in D} 1 p(x) = 1$$

That is, $M_X(0)$ is the sum of all the probabilities, so it must always be 1. However, in order for the mgf to be useful in generating moments, it will need to be defined for an interval of values of t including 0 in its interior. The moment generating function fails to exist in cases when moments themselves fail to exist (see Example 2.49 below).

Example 2.46 The simplest example of an mgf is for a Bernoulli distribution, where only the X values 0 and 1 receive positive probability. Let X be a Bernoulli random variable with $p(0) = 1/3$ and $p(1) = 2/3$. Then

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x \in D} e^{tx} p(x) = e^{t \cdot 0} \cdot (1/3) + e^{t \cdot 1} \cdot (2/3) = (1/3) + (2/3)e^t$$

A Bernoulli random variable will always have an mgf of the form $p(0) + p(1)e^t$, a well-defined function for all values of t.   ∎

A key property of the mgf is its "uniqueness," the fact that it completely characterizes the underlying distribution.

MGF UNIQUENESS THEOREM

If the mgf exists and is the same for two distributions, then the two distributions are identical. That is, the moment generating function uniquely specifies the probability distribution; there is a one-to-one correspondence between distributions and mgfs.

The proof of this theorem, originally due to Laplace, requires some sophisticated mathematics and is beyond the scope of this textbook.

Example 2.47 Let X, the number of claims submitted on a renter's insurance policy on a given year, have mgf $M_X(t) = .7 + .2e^t + .1e^{2t}$. It follows that X must have the pmf $p(0) = .7$, $p(1) = .2$, and $p(2) = .1$—because if we use this pmf to obtain the mgf, we get $M_X(t)$, and the distribution is uniquely determined by its mgf. ∎

Example 2.48 Consider testing individuals' blood samples one by one in order to find someone whose blood type is Rh+. Suppose X, the number of tested samples, has a geometric distribution with $p = .85$:

$$p(x) = .85(.15)^{x-1} \quad \text{for } x = 1, 2, 3, ...$$

Determining the moment generating function here requires using the formula for the sum of a geometric series: $1 + r + r^2 + \cdots = 1/(1 - r)$ for $|r| < 1$. The moment generating function is

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} p(x) = \sum_{x=1}^{\infty} e^{tx} .85(.15)^{x-1} = .85e^t \sum_{x=1}^{\infty} e^{t(x-1)}(.15)^{x-1}$$

$$= .85e^t \sum_{x=1}^{\infty} (.15e^t)^{x-1} = .85e^t \left[1 + .15e^t + (.15e^t)^2 + \cdots\right] = \frac{.85e^t}{1 - .15e^t}$$

The condition on r requires $|.15e^t| < 1$. Dividing by .15 and taking logs, this gives $t < -\ln(.15) \approx 1.90$, i.e., this function is defined in the interval $(-\infty, 1.90)$. The result is an interval of values that includes 0 in its interior, so the mgf exists. As a check, $M_X(0) = .85/(1 - .15) = 1$, as required. ∎

Example 2.49 Reconsider Example 2.20, where $p(x) = k/x^2$, $x = 1, 2, 3, \ldots$. Recall that $E(X)$ does not exist for this distribution, portending a problem for the existence of the mgf:

$$M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} \frac{k}{x^2}$$

With the help of tests for convergence such as the ratio test, we find that the series converges if and only if $e^t \le 1$, which means that $t \le 0$, i.e., the mgf is only defined on the interval $(-\infty, 0]$. Because zero is on the *boundary* of this interval, not the interior of the interval (the interval must include both positive and negative values), the mgf of this distribution does not exist. In any case, it could not be useful for finding moments, because X does not have even a first moment (mean). ∎

2.7.2   Obtaining Moments from the MGF

We now turn to the computation of moments from the mgf. For any positive integer r, let $M_X^{(r)}(t)$ denote the rth derivative of $M_X(t)$. By computing this and then setting $t = 0$, we get the rth moment about 0.

THEOREM

If the mgf of X exists, then $E(X^r)$ is finite for all positive integers r, and

$$E(X^r) = M_X^{(r)}(0) \tag{2.21}$$

Proof   The proof of the existence of all moments is beyond the scope of this book. We will show that Eq. (2.21) is true for $r = 1$ and $r = 2$. A proof by mathematical induction can be used for general r. Differentiate:

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} \sum_{x \in D} e^{xt} p(x) = \sum_{x \in D} \frac{d}{dt} e^{xt} p(x) = \sum_{x \in D} x e^{xt} p(x)$$

where we have interchanged the order of summation and differentiation. (This is justified inside the interval of convergence, which includes 0 in its interior.) Next set $t = 0$ to obtain the first moment:

$$M_X'(0) = M_X^{(1)}(0) = \sum_{x \in D} x e^{x(0)} p(x) = \sum_{x \in D} x p(x) = E(X)$$

Differentiating a second time gives

$$\frac{d^2}{dt^2} M_X(t) = \frac{d}{dt} \sum_{x \in D} x e^{xt} p(x) = \sum_{x \in D} x \frac{d}{dt} e^{xt} p(x) = \sum_{x \in D} x^2 e^{xt} p(x)$$

Set $t = 0$ to get the second moment:

$$M_X''(0) = M_X^{(2)}(0) = \sum_{x \in D} x^2 p(x) = E(X^2) \qquad \blacksquare$$

For the pmfs in Examples 2.45 and 2.46, this may seem like needless work—after all, for simple distributions with just a few values, we can quickly determine the mean, variance, etc. The real utility of the mgf arises for more complicated distributions.

Example 2.50   (Example 2.48 continued) Recall that $p = .85$ is the probability of a person having Rh + blood and we keep checking people until we find one with this blood type. If X is the number of people we need to check, then $p(x) = .85(.15)^{x-1}$, $x = 1, 2, 3, \ldots$, and the mgf is

$$M_X(t) = E\left(e^{tX}\right) = \frac{.85e^t}{1 - .15e^t}$$

Differentiating with the help of the quotient rule,

$$M_X'(t) = \frac{.85e^t}{(1 - .15e^t)^2}$$

Setting $t = 0$ then gives $\mu = E(X) = M_X'(0) = 1/.85 = 1.176$. This corresponds to the formula $1/p$ for a geometric distribution.

To get the second moment, differentiate again:

$$M_X''(t) = \frac{.85e^t(1 + .15e^t)}{(1 - .15e^t)^3}$$

Setting $t = 0$, $E(X^2) = M_X''(0) = \frac{1.15}{.85^2}$. Now use the variance shortcut formula:

$$\text{Var}(X) = \sigma^2 = E(X^2) - \mu^2 = \frac{1.15}{.85^2} - \left(\frac{1}{.85}\right)^2 = \frac{.15}{.85^2} = .2076$$

This matches the variance formula $(1 - p)/p^2$ given without proof toward the end of Sect. 2.6. ∎

As mentioned in Sect. 2.3, it is common to transform a rv X using a linear function $Y = aX + b$. What happens to the mgf when we do this?

> **PROPOSITION**
> Let X have the mgf $M_X(t)$ and let $Y = aX + b$. Then $M_Y(t) = e^{bt}M_X(at)$.

Example 2.51 Let X be a Bernoulli random variable with $p(0) = 20/38$ and $p(1) = 18/38$. Think of X as the number of wins, 0 or 1, in a single play of roulette. If you play roulette at an American casino and bet on red, then your chances of winning are $18/38$ because 18 of the 38 possible outcomes are red. From Example 2.46, $M_X(t) = 20/38 + e^t(18/38)$. Suppose you bet \$5 on red, and let Y be your winnings. If $X = 0$ then $Y = -5$, and if $X = 1$ then $Y = 5$. The linear equation $Y = 10X - 5$ gives the appropriate relationship.

This equation is of the form $Y = aX + b$ with $a = 10$ and $b = -5$, so by the foregoing proposition

$$M_Y(t) = e^{bt}M_X(at) = e^{-5t}M_X(10t)$$
$$= e^{-5t}\left[\frac{20}{38} + e^{10t}\frac{18}{38}\right] = e^{-5t} \cdot \frac{20}{38} + e^{5t} \cdot \frac{18}{38}$$

This implies that the pmf of Y is $p(-5) = 20/38$ and $p(5) = 18/38$; moreover, we can compute the mean (and other moments) of Y directly from this mgf. ∎

2.7.3   MGFs of Common Distributions

Several of the distributions presented in this chapter (binomial, Poisson, negative binomial) have fairly simple expressions for their moment generating functions. These mgfs, in turn, allow us to determine the means and variances of the distributions without some rather unpleasant summation. (Additionally, we will use these mgfs to prove some more advanced distributional properties in Chap. 4.)

To start, determining the moment generating function of a binomial rv requires use of the binomial theorem: $(a + b)^n = \sum_{x=0}^n \binom{n}{x}a^x b^{n-x}$. Then

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} b(x; n, p) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

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

The mean and variance can be obtained by differentiating $M_X(t)$:

$$M_X'(t) = n(pe^t + 1 - p)^{n-1} pe^t \quad \Rightarrow \quad \mu = M_X'(0) = np;$$

$$M_X''(t) = n(n-1)(pe^t + 1 - p)^{n-2} pe^t pe^t + n(pe^t + 1 - p)^{n-1} pe^t \Rightarrow$$

$$E(X^2) = M_X''(0) = n(n-1)p^2 + np \Rightarrow$$

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

$$= n(n-1)p^2 + np - n^2 p^2 = np - np^2 = np(1-p),$$

in accord with the proposition in Sect. 2.4.

Derivation of the Poisson mgf utilizes the series expansion $\sum_{x=0}^{\infty} u^x / x! = e^u$:

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{-\mu} e^{\mu e^t} = e^{\mu(e^t - 1)}$$

Successive differentiation then gives the mean and variance identified in Sect. 2.5 (see Exercise 127).

Finally, derivation of the negative binomial mgf is based on Newton's generalization of the binomial theorem. The result (see Exercise 124) is

$$M_X(t) = \left(\frac{pe^t}{1 - (1-p)e^t} \right)^r$$

The geometric mgf is just the special case $r = 1$ (cf. Example 2.48 above). There is unfortunately no simple expression for the mgf of a hypergeometric rv.

2.7.4   Exercises: Section 2.7 (107–128)

107. For the entry-level employees of a certain fast food chain, the pmf of $X =$ highest grade level completed is specified by $p(9) = .01$, $p(10) = .05$, $p(11) = .16$, and $p(12) = .78$.
    (a) Determine the moment generating function of this distribution.
    (b) Use (a) to find $E(X)$ and SD(X).
108. For a new car the number of defects X has the distribution given by the accompanying table. Find $M_X(t)$ and use it to find $E(X)$ and SD(X).

x	0	1	2	3	4	5	6
$p(x)$	.04	.20	.34	.20	.15	.04	.03

109. In flipping a fair coin let X be the number of tosses to get the first head. Then $p(x) = .5^x$ for $x = 1, 2, 3, \ldots$. Find $M_X(t)$ and use it to get $E(X)$ and SD(X).
110. If you toss a fair die with outcome X, $p(x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$. Determine $M_X(t)$.
111. Find the skewness coefficients of the distributions in the previous four exercises. Do these agree with the "shape" of each distribution?

112. Given $M_X(t) = .2 + .3e^t + .5e^{3t}$, find $p(x)$, $E(X)$, $\text{Var}(X)$.
113. If $M_X(t) = 1/(1 - t^2)$, find $E(X)$ and $\text{Var}(X)$.
114. Show that $g(t) = te^t$ cannot be a moment generating function.
115. Using a calculation similar to the one in Example 2.48 show that, if X has a geometric distribution with parameter p, then its mgf is

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

Assuming that Y has mgf $M_Y(t) = .75e^t/(1 - .25e^t)$, determine the probability mass function $p(y)$ with the help of the uniqueness property.
116. (a) Prove the result in the second proposition: $M_{aX+b}(t) = e^{bt}M_X(at)$.
    (b) Let $Y = aX + b$. Use (a) to establish the relationships between the means and variances of X and Y.
117. Let $M_X(t) = e^{5t+2t^2}$ and let $Y = (X - 5)/2$. Find $M_Y(t)$ and use it to find $E(Y)$ and $\text{Var}(Y)$.
118. Let X have the moment generating function of Example 2.48 and let $Y = X - 1$. Recall that X is the number of people who need to be checked to get someone who is Rh+, so Y is the number of people checked before the first Rh+ person is found. Find $M_Y(t)$.
119. Let X be the number of points earned by a randomly selected student on a 10 point quiz, with possible values 0, 1, 2, …, 10 and pmf $p(x)$, and suppose the distribution has a skewness coefficient of c. Now consider reversing the probabilities in the distribution, so that $p(0)$ is interchanged with $p(10)$, $p(1)$ is interchanged with $p(9)$, and so on. Show that the skewness coefficient of the resulting distribution is $-c$. [*Hint*: Let $Y = 10 - X$ and show that Y has the reversed distribution. Use this fact to determine μ_Y and then the value of skewness coefficient for the Y distribution.]
120. Let $M_X(t)$ be the moment generating function of a rv X, and define a new function by

$$L_X(t) = \ln[M_X(t)]$$

Show that (a) $L_X(0) = 0$, (b) $L'_X(0) = \mu$, and (c) $L''_X(0) = \sigma^2$.
121. Refer back to Exercise 120. If $M_X(t) = e^{5t+2t^2}$ then find $E(X)$ and $\text{Var}(X)$ by differentiating
    (a) $M_X(t)$
    (b) $L_X(t)$
122. Refer back to Exercise 120. If $M_X(t) = e^{5(e^t-1)}$ then find $E(X)$ and $\text{Var}(X)$ by differentiating
    (a) $M_X(t)$
    (b) $L_X(t)$
123. Obtain the moment generating function of the number of failures, $n - X$, in a binomial experiment, and use it to determine the expected number of failures and the variance of the number of failures. Are the expected value and variance intuitively consistent with the expressions for $E(X)$ and $\text{Var}(X)$? Explain.
124. Newton's generalization of the binomial theorem can be used to show that, for any positive integer r,

$$(1 - u)^{-r} = \sum_{k=0}^{\infty} \binom{r + k - 1}{r - 1} u^k$$

Use this to derive the negative binomial mgf presented in this section. Then obtain the mean and variance of a binomial rv using this mgf.

125. If X is a negative binomial rv, then $Y = X - r$ is the number of failures preceding the rth success. Obtain the mgf of Y and then its mean value and variance.
126. Refer back to Exercise 120. Obtain the negative binomial mean and variance from $L_X(t) = \ln[M_X(t)]$.
127. (a)   Use derivatives of $M_X(t)$ to obtain the mean and variance for the Poisson distribution.
     (b)   Obtain the Poisson mean and variance from $L_X(t) = \ln[M_X(t)]$. In terms of effort, how does this method compare with the one in part (a)?
128. Show that the binomial moment generating function converges to the Poisson moment generating function if we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. [*Hint*: Use the calculus theorem that was used in showing that the binomial pmf converges to the Poisson pmf.] There is, in fact, a theorem saying that convergence of the mgf implies convergence of the probability distribution. In particular, convergence of the binomial mgf to the Poisson mgf implies $b(x; n, p) \to p(x; \mu)$.

2.8   Simulation of Discrete Random Variables

Probability calculations for complex systems often depend on the behavior of various random variables. When such calculations are difficult or impossible, simulation is the fallback strategy. In this section, we give a general method for simulating an arbitrary discrete random variable and consider implementations in existing software for simulating common discrete distributions.

Example 2.52 Refer back to the distribution of Example 2.11 for the random variable $X = $ the amount of memory (GB) in a purchased flash drive, and suppose we wish to simulate X. Recall from Sect. 1.6 that we begin with a "standard uniform" random number generator, i.e., a software function that generates evenly distributed numbers in the interval [0, 1). Our goal is to convert these decimals into the values of X with the probabilities specified by its pmf: 5% 1s, 10% 2s, 35% 4s, and so on. To that end, we partition the interval [0, 1) according to these percentages: [0, .05) has probability .05; [.05, .15) has probability .1, since the length of the interval is .1; [.15, .50) has probability $.50 - .15 = .35$; etc. Proceed as follows: given a value u from the RNG,

- If $0 \le u < .05$, assign the value 1 to the variable x.
- If $.05 \le u < .15$, assign $x = 2$.
- If $.15 \le u < .50$, assign $x = 4$.
- If $.50 \le u < .90$, assign $x = 8$.
- If $.90 \le u < 1$, assign $x = 16$.

Repeating this algorithm n times gives n simulated values of X. Programs in Matlab and R that implement this algorithm appear in Fig. 2.10; both return a vector, x, containing $n = 10,000$ simulated values of the specified distribution.

Figure 2.11 shows a graph of the results of executing the code, in the form of a *histogram*: the height of each rectangle corresponds to the relative frequency of each x value in the simulation (i.e., the number of times that value occurred, divided by 10,000). The exact pmf of X is superimposed for comparison; as expected, simulation results are similar, but not identical, to the theoretical distribution.

a
```
x=zeros(10000,1);
for i=1:10000
    u=rand;
    if u<.05
        x(i)=1;
    elseif u<.15
        x(i)=2;
    elseif u<.50
        x(i)=4;
    elseif u<.90
        x(i)=8;
    else
        x(i)=16;
    end
end
```

b
```
x <- NULL
for (i in 1:10000){
    u=runif(1)
    if (u<.05)
        x[i]<-1
    else if (u<.15)
        x[i]<-2
    else if (u<.50)
        x[i]<-4
    else if (u<.90)
        x[i]<-8
    else
        x[i]<-16
}
```

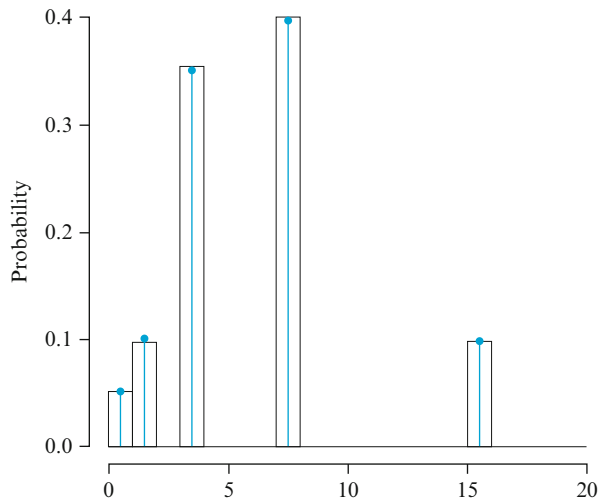**Fig. 2.10** Simulation code: (**a**) Matlab; (**b**) R

Fig. 2.11 Simulation and exact distribution for Example 2.52

Later in this section, we will present a faster, built-in way to simulate discrete distributions in Matlab and R. The method introduced here will, however, prove useful in adapting to the case of continuous random variables in Chap. 3.   ∎

In the preceding example, the selected subintervals of [0, 1) were not our only choices—any five intervals with lengths .05, .10, .35, .40, and .10 would produce the desired result. However, those particular five subintervals have one desirable feature: the "cut points" for the intervals (i.e., 0, .05, .15, .50, .90, and 1) are precisely the possible heights of the graph of the cdf, $F(x)$. This permits a geometric interpretation of the algorithm, which can be seen in Fig. 2.12. The value u provided by the RNG corresponds to a position on the vertical axis between 0 and 1; we then "invert" the cdf by matching this u-value back to one of the gaps in the graph of $F(x)$, denoted by dashed lines in Fig. 2.12. If the gap occurs at horizontal position x, then x is our simulated value of the rv X for that run of the simulation. This is often referred to as the **inverse cdf method** for simulating discrete random variables. The general method is spelled out in the accompanying box.

Fig. 2.12 The inverse cdf method for Example 2.52
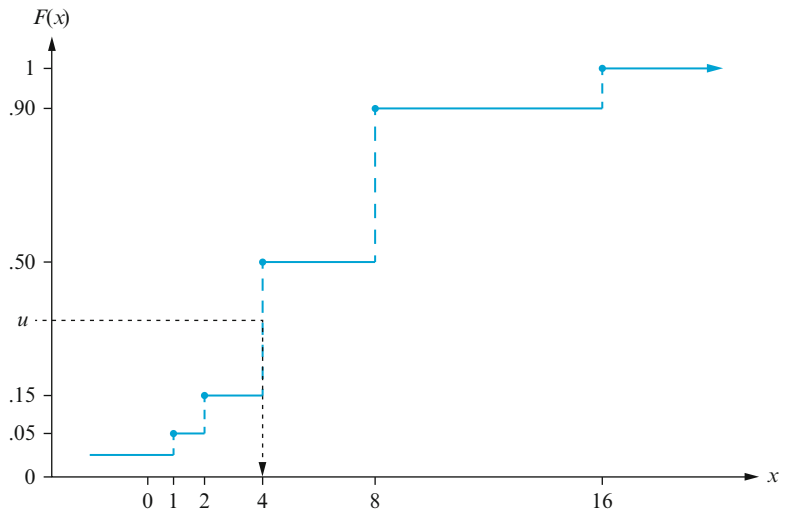
Inverse cdf Method for Simulating Discrete Random Variables

Let X be a discrete random variable taking on values $x_1 < x_2 < \ldots$ with corresponding probabilities $p_1, p_2, \ldots$. Define $F_0 = 0$; $F_1 = F(x_1) = p_1$; $F_2 = F(x_2) = p_1 + p_2$; and, in general, $F_k = F(x_k) = p_1 + \cdots + p_k = F_{k-1} + p_k$. To simulate a value of X, proceed as follows:

1. Use an RNG to produce a value, u, from $[0, 1)$.
2. If $F_{k-1} \le u < F_k$, then assign $x = x_k$.

Example 2.53 (Example 2.52 continued): Suppose the prices for the flash drives, in increasing order of memory size, are \$10, \$15, \$20, \$25, and \$30. If the store sells 80 flash drives in a week, what's the probability they will make a gross profit of at least \$1800?

Let $Y =$ the amount spent on a flash drive, which has the following pmf:

y	10	15	20	25	30
$p(y)$	.05	.10	.35	.40	.10

The gross profit for 80 purchases is the sum of 80 values from this distribution. Let $A = \{$gross profit $\ge \$1800\}$. We can use simulation to estimate $P(A)$, as follows:

0. Set a counter for the number of times A occurs to zero.

   Repeat n times:

1. Simulate 80 values $y_1, \ldots, y_{80}$ from the above pmf (using for example an inverse cdf program similar to those displayed in Fig. 2.10).
2. Compute the week's gross profit, $g = y_1 + \cdots + y_{80}$.
3. If $g \ge 1800$, add 1 to the count of occurrences for A.

Once the n runs are complete, then $\hat{P}(A) = ($count of the occurrences of $A)/n$.

Figure 2.13 shows the resulting values of g for $n = 10{,}000$ simulations in R. In effect, our program is simulating a random variable $G = Y_1 + \ldots + Y_{80}$ whose pmf is not known (in light of all the possible
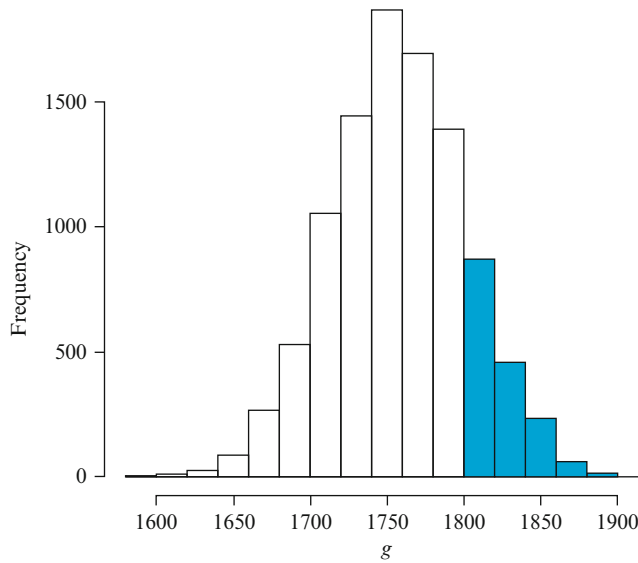
Fig. 2.13 Simulated distribution of weekly gross profit for Example 2.53

G values, it would not be worthwhile to attempt to determine its pmf analytically). The highlighted bars in Fig. 2.13 correspond to g values of at least \$1800; in our simulation, such values occurred 1940 times. Thus, $\hat{P}(A) = 1940/10,000 = .194$, with an estimated standard error of $\sqrt{.194(1 - .194)/10,000} = .004$. ∎

2.8.1 Simulations Implemented in R and Matlab

Earlier in this section, we presented the inverse cdf method as a general way to simulate discrete distributions applicable in any software. In fact, one can simulate generic discrete rvs in both Matlab and R by clever use of the built-in `randsample` and `sample` functions, respectively. We saw these functions in the context of probability simulation in Chap. 1. Both are designed to generate a random sample from any selected set of values (even including text values, if desired); the "clever" part is that both can accommodate a set of weights. The following short example illustrates their use.

To simulate, say, 35 values from the pmf in Example 2.52, one can use the following code in Matlab:

```
randsample([1,2,4,8,16],35,true,[.05, .10, .35, .40, .10])
```

The function takes four arguments: the list of *x*-values, the desired number of simulated values (the "sample size"), whether to sample with replacement (here, `true`), and the list of probabilities in the same order as the *x*-values. The corresponding call in R is

```
sample(c(1,2,4,8,16),35,TRUE,c(.05, .10, .35, .40, .10))
```

Thanks to the ubiquity of the binomial, Poisson, and other distributions in probability modeling, many software packages have built-in tools for simulating values from these distributions. Table 2.6 summarizes the relevant functions in Matlab and R; the input argument *size* refers to the desired number of simulated values of the distribution.

Table 2.6 Functions to simulate major discrete distributions in Matlab and R

Distribution	Matlab code	R code
Binomial	`binornd(`$n, p, size, 1$`)`	`rbinom(`$size, n, p$`)`
Poisson	`poissrnd(`$\mu, size, 1$`)`	`rpois(`$size, \mu$`)`
Hypergeometric	`hygernd(`$N, M, n, size, 1$`)`	`rhyper(`$size, M, N - M, n$`)`
Negative binomial	`nbinrnd(`$r, p, size, 1$`)`	`rnbinom(`$size, r, p$`)`

A word of warning (really, a reminder) about the way software treats the negative binomial distribution: both Matlab and R define a negative binomial rv as the number of failures preceding the rth success, which differs from our definition. Assuming you want to simulate the number of *trials* required to achieve r successes, execute the code in the last line of Table 2.6 and then add r to each value.

Example 2.54 The number of customers shipping express mail packages at a certain store during any particular hour of the day is a Poisson rv with mean 5. Each such customer has 1, 2, 3, or 4 packages with probabilities .4, .3, .2, and .1, respectively. Let's carry out a simulation to estimate the probability that at most 10 packages are shipped during any particular hour.

Define an event $A = \{$at most 10 packages shipped in an hour$\}$. Our simulation to estimate $P(A)$ proceeds as follows.

0. Set a counter for the number of times A occurs to zero.

   Repeat n times:

1. Simulate the number of customers in an hour, C, which is Poisson with $\mu = 5$.
2. For each of the C customers, simulate the number of packages shipped according to the pmf above.
3. If the total number of packages shipped is at most 10, add 1 to the counter for A.

   Matlab and R code to implement this simulation appear in Fig. 2.14.

a
```
A=0;
for i=1:10000
    c=poissrnd(5,1);
    packages = randsample([1,2,3,4],c,
            true,[.4,.3,.2,.1]);
    if sum(packages)<=10
        A=A+1;
    end
end
```

b
```
A <- 0
for (i in 1:10000){
    c<-rpois(1,5)
    packages <- sample(c(1,2,3,4),c,
            TRUE,c(.4,.3,.2,.1))
    if (sum(packages)<=10){
        A<-A+1
    }
}
```

Fig. 2.14 Simulation code for Example 2.54: (**a**) Matlab; (**b**) R

In Matlab, 10,000 simulations resulted in 10 or fewer packages 5752 times, for an estimated probability of $\hat{P}(A) = .5752$, with an estimated standard error of $\sqrt{.5752(1 - .5752)/10{,}000} = .0049$.   ∎

2.8.2 Simulation Mean, Standard Deviation, and Precision

In Sect. 1.6 and in the preceding examples, we used simulation to estimate the probability of an event. But consider the "gross profit" variable G in Example 2.53: since we have 10,000 simulated values of this variable, we should be able to estimate its mean μ_G and its standard deviation σ_G. More generally,

suppose we have simulated n values $x_1, \ldots, x_n$ of a random variable X. Then the following quantities based on our observed values serve as suitable estimates.

DEFINITION

For a set of numerical values $x_1, \ldots, x_n$, the **sample mean**, denoted by $\bar{x}$, is

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The **sample standard deviation** of these numerical values, denoted by s, is

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

If $x_1, \ldots, x_n$ represent simulated values of a random variable X, then we may estimate the expected value and standard deviation of X by $\hat{\mu} = \bar{x}$ and $\hat{\sigma} = s$, respectively.

The justification for the use of the divisor $n - 1$ in s will be discussed in Chap. 5.

In Sect. 1.6, we introduced the standard error of an estimated probability, which quantifies the precision of a simulation result $\hat{P}(A)$ as an estimate of a "true" probability $P(A)$. By analogy, it is possible to quantify the amount by which a sample mean, $\bar{x}$, will generally differ from the corresponding expected value μ. For n simulated values of a random variable, with sample standard deviation s, the (**estimated**) **standard error of the mean** is

$$\frac{s}{\sqrt{n}} \tag{2.22}$$

Expression (2.22) will be derived in Chap. 4. As with an estimated probability, the formula indicates that the precision of $\bar{x}$ increases (i.e., its standard error *decreases*) as n increases, but not very quickly. To increase the precision of $\bar{x}$ as an estimate of μ by a factor of 10 (one decimal place) requires increasing the number of simulation runs, n, by a factor of 100. Unfortunately, there is no general formula for the standard error of s as an estimate of σ.

Example 2.55 (Ex. 2.53 continued) The 10,000 simulated values of the random variable G, which we denote by $g_1, \ldots, g_{10000}$, are displayed in the histogram in Fig. 2.13. From these simulated values, we can estimate both the expected value and standard deviation of G:

$$\hat{\mu}_G = \bar{g} = \frac{1}{10,000} \sum_{i=1}^{10,000} g_i = 1759.62$$

$$\hat{\sigma}_G = s = \sqrt{\frac{1}{10,000-1} \sum_{i=1}^{10,000} (g_i - \bar{g})^2} = \sqrt{\frac{1}{9999} \sum_{i=1}^{10,000} (g_i - 1759.62)^2} = 43.50$$

We estimate that the average weekly gross profit from flash drive sales is \$1759.62, with a standard deviation of \$43.50. Neither of these computations was performed by hand, of course: if the n simulated values of a variable are stored in a vector x, then mean(x) and sd(x) in R will provide the sample mean and standard deviation, respectively. In Matlab, the calls are mean(x) and std(x).

Applying Eq. (2.22), the (estimated) standard error of $\bar{g}$ is $s/\sqrt{n} = 43.50/\sqrt{10,000} = 0.435$. If 10,000 runs are used to simulate G, it's estimated that the resulting sample mean will differ from

$E(G)$ by roughly 0.435. (In contrast, the sample standard deviation, s, estimates that the gross profit for a single week—i.e., a single observation g—typically differs from $E(G)$ by about \$43.50.) ∎

In Chap. 4, we will see how the expected value and variance of random variables like G, that are sums of a fixed number of other rvs, can be obtained analytically.

Example 2.56 The "help desk" at a university's computer center receives both hardware and software queries. Let X and Y be the number of hardware and software queries, respectively, in a given day. Each can be modeled by a Poisson distribution with mean 20. Because computer center employees need to be allocated efficiently, of interest is the *difference* between the sizes of the two queues: $D = |X - Y|$. Let's use simulation to estimate (1) the probability the queue sizes differ by more than 5; (2) the expected difference; (3) the standard deviation of the difference.

Figure 2.15 shows Matlab and R code to simulate this process. In both languages, the code exploits the built-in Poisson simulator, as well as the fact that 10,000 simulated values may be called simultaneously.

a
```
X=poissrnd(20,10000,1);
Y=poissrnd(20,10000,1);
D=abs(X-Y);
sum((D>5))
mean(D)
std(D)
```

b
```
X<-rpois(10000,20)
Y<-rpois(10000,20)
D<-abs(X-Y)
sum((D>5))
mean(D)
sd(D)
```

Fig. 2.15 Simulation code for Example 2.56: (**a**) Matlab; (**b**) R

The line `sum((D>5))` performs two operations: first, `(D>5)` determines if each simulated d value exceeds 5, returning a logical vector of bits; second, `sum()` tallies the "success" bits (1s or TRUEs) and gives a count of the number of times the event $\{D > 5\}$ occurred in the 10,000 simulations. The results from one run in Matlab were

$$\hat{P}(D > 5) = \frac{3843}{10,000} = .3843 \quad \hat{\mu}_D = \bar{d} = 5.0380 \quad \hat{\sigma}_D = s = 3.8436$$

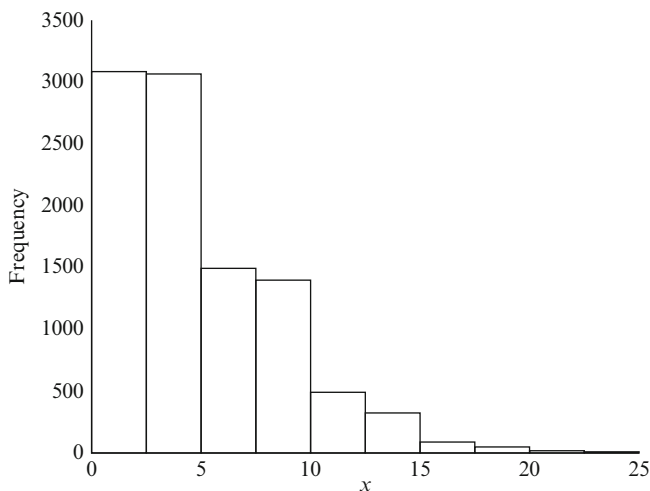A histogram of the simulated values of D appears in Fig. 2.16.

Fig. 2.16 Simulation histogram of D in Example 2.56   ∎

2.8.3   Exercises: Section 2.8 (129–141)

129. Consider the pmf given in Exercise 30 for the random variable $Y =$ the number of moving violations for which the a randomly selected insured individual was cited during the last 3 years. Write a program to simulate this random variable, then use your simulation to estimate $E(Y)$ and $SD(Y)$. How do these compare to the exact values of $E(Y)$ and $SD(Y)$?

130. Consider the pmf given in Exercise 32 for the random variable $X =$ capacity of a purchased freezer. Write a program to simulate this random variable, then use your simulation to estimate $E(X)$ and $SD(X)$. How do these compare to the exact values of $E(X)$ and $SD(X)$?

131. Suppose person after person is tested for the presence of a certain characteristic. The probability that any individual tests positive is .75. Let $X =$ the number of people who must be tested to obtain five *consecutive* positive test results. Use simulation to estimate $P(X \le 25)$.

132. *The matching problem.* Suppose that N items labeled 1, 2, …, N are shuffled so that they are in random order. Of interest is how many of these will be in their "correct" positions (e.g., item #5 situated at the 5th position in the sequence, etc.) after shuffling.
    (a) Write a program that simulates a permutation of the numbers 1 to N and then records the value of the variable $X =$ number of items in the correct position.
    (b) Set $N = 5$ in your program, and use at least 10,000 simulations to estimate $E(X)$, the expected number of items in the correct position.
    (c) Set $N = 52$ in your program (as if you were shuffling a deck of cards), and use at least 10,000 simulations to estimate $E(X)$. What do you discover? Is this surprising?

133. Exercise 109 of Chap. 1 referred to a multiple-choice exam in which 10 of the questions have two options, 13 have three options, 13 have four options, and the other 4 have five options. Let $X =$ the number of questions a student gets right, assuming s/he is completely guessing.
    (a) Write a program to simulate X, and use your program to estimate the mean and standard deviation of X.
    (b) Estimate the probability a student will score at least one standard deviation above the mean.

134. Example 2.53 of this section considered the gross profit G resulting from selling flash drives to 80 customers per week. Of course, it isn't realistic for the number of customers to remain fixed from week to week. So, instead, imagine the number of customers buying flash drives in a week follows a Poisson distribution with mean 80, and that the amount paid by each customer follows the distribution for Y provided in that example. Write a program to simulate the random variable G, and use your simulation to estimate
    (a) The probability that weekly gross sales are at least $1,800.
    (b) The mean of G.
    (c) The standard deviation of G.

135. Exercise 21 (Sect. 2.2) investigated Benford's law, a discrete distribution with pmf given by $p(x) = \log_{10}((x + 1)/x)$ for $x = 1, 2, …, 9$. Use the inverse cdf method to write a program that simulates the Benford's law distribution. Then use your program to estimate the expected value and variance of this distribution.

136. Recall that a geometric rv has pmf $p(x) = p(1 - p)^{x-1}$ for $x = 1, 2, 3, ….$. In Example 2.12, it was shown that the cdf of this distribution is given by $F(x) = 1 - (1 - p)^x$ for positive integers x.
    (a) Write a program that implements the inverse cdf method to simulate a geometric distribution. Your program should have as inputs the numerical value of p and the desired sample size.
    (b) Use your program to simulate 10,000 values from a geometric rv X with $p = .85$. From these values, estimate each of the following: $P(X \le 2)$, $E(X)$, $SD(X)$. How do these compare to the corresponding exact values?

137. Tickets for a particular flight are $250 apiece. The plane seats 120 passengers, but the airline will knowingly overbook (i.e., sell more than 120 tickets), because not every paid passenger shows up. Let t denote the number of tickets the airline sells for this flight, and assume the number of passengers that actually show up for the flight, X, follows a Bin(t, .85) distribution. Let $B =$ the number of paid passengers who show up at the airport but are denied a seat on the plane, so $B = X - 120$ if $X > 120$ and $B = 0$ otherwise. If the airline must compensate these passengers with $500 apiece, then the profit the airline makes on this flight is $250t - 500B$. (Notice t is fixed, but B is random.)

    (a) Write a program to simulate this scenario. Specifically, your program should take in t as an input and return many values of the profit variable $250t - 500B$.

    (b) The airline wishes to determine the optimal value of t, i.e., the number of tickets to sell that will maximize their expected profit. Run your program for $t = 140, 141, \ldots, 150$, and record the average profit from many runs under each of these settings. What value of t appears to return the largest value? [*Note:* If a clear winner does not emerge, you might need to increase the number of runs for each t value!]

138. Imagine the following simple game: flip a fair coin repeatedly, winning $1 for every head and losing $1 for every tail. Your net winnings will potentially oscillate between positive and negative numbers as play continues. How many times do you think net winnings will *change signs* in, say, 1000 coin flips? 5000 flips?

    (a) Let $X =$ the number of sign changes in 1000 coin flips. Write a program to simulate X, and use your program to estimate the probability of at least 10 sign changes.

    (b) Use your program to estimate $E(X)$ and $SD(X)$. Does your estimate for $E(X)$ match your intuition for the number of sign changes?

    (c) Repeat parts (a)–(b) with 5000 flips.

139. Exercise 39 (Sect. 2.3) describes the game Plinko from *The Price is Right*. Each contestant drops between one and 5 chips down the Plinko board, depending on how well s/he prices several small items. Suppose the random variable $C =$ number of chips earned by a contestant has the following distribution:

c	1	2	3	4	5
$p(c)$	.03	.15	.35	.34	.13

The winnings from each chip follow the distribution presented in Exercise 39. Write a program to simulate Plinko; you will need to consider both the number of chips a contestant earns and how much money is won on each of those chips. Use your simulation estimate the answers to the following questions:

    (a) What is the probability a contestant wins more than $11,000?

    (b) What is a contestant's expected winnings?

    (c) What is the corresponding standard deviation?

    (d) In fact, a player gets one Plinko chip for free and can earn the other four by guessing the prices of small items (waffle irons, alarm clocks, etc.). Assume the player has a 50–50 chance of getting each price correct, so we may write $C = 1 + R$, where $R \sim$ Bin(4, .5). Use this revised model for C to estimate the answers to (a)–(c).

140. Recall the *Coupon Collector's Problem* described in the book's Introduction and again in Exercise 114 of Chap. 1. Let $X =$ the number of cereal boxes purchased in order to obtain all 10 coupons.

    (a) Use a simulation program to estimate $E(X)$ and $SD(X)$. Also compute the estimated standard error of your sample mean.

(b) How does your estimate of $E(X)$ compare to the theoretical answer given in the Introduction?

(c) Repeat (a) with 20 coupons required instead of 10. Does it appear to take roughly twice as long to collect 20 coupons as 10? More than twice as long? Less?

141. A small high school holds its graduation ceremony in the gym. Because of seating constraints, students are limited to a maximum of four tickets to graduation for family and friends. Suppose 30% of students want four tickets, 25% want three, 25% want two, 15% want one, and 5% want none.

(a) Write a simulation for 150 graduates requesting tickets, where students' requests follow the distribution described above. In particular, keep track of the variable T = the total number of tickets requested by these 150 students.

(b) The gym can seat a maximum of 410 guests. Based on your simulation, estimate the probability that all students' requests can be accommodated.

2.9   Supplementary Exercises (142–170)

142. Consider a deck consisting of seven cards, marked 1, 2, . . ., 7. Three of these cards are selected at random. Define an rv W by W = the sum of the resulting numbers, and compute the pmf of W. Then compute $E(W)$ and $\text{Var}(W)$. [*Hint*: Consider outcomes as unordered, so that (1, 3, 7) and (3, 1, 7) are not different outcomes. Then there are 35 outcomes, and they can be listed.] (This type of rv actually arises in connection with *Wilcoxon's rank-sum test*, in which there is an x sample and a y sample and W is the sum of the ranks of the x's in the combined sample.)

143. After shuffling a deck of 52 cards, a dealer deals out 5. Let X = the number of suits represented in the five-card hand.

(a) Show that the pmf of X is

x	1	2	3	4
$p(x)$	.002	.146	.588	.264

[*Hint*: $p(1) = 4P$(all are spades), $p(2) = 6P$(only spades and hearts with at least one of each), and $p(4) = 4P(2 \text{ spades} \cap \text{one of each other suit})$.]

(b) Compute $E(X)$ and $\text{SD}(X)$.

144. The negative binomial rv X was defined as the number of trials necessary to obtain the rth S. Let Y = the number of F's preceding the rth S. In the same manner in which the pmf of X was derived, derive the pmf of Y.

145. Of all customers purchasing automatic garage-door openers, 75% purchase a chain-driven model. Let X = the number among the next 15 purchasers who select the chain-driven model.

(a) What is the pmf of X?

(b) Compute $P(X > 10)$.

(c) Compute $P(6 \le X \le 10)$.

(d) Compute $E(X)$ and $\text{SD}(X)$.

(e) If the store currently has in stock 10 chain-driven models and 8 shaft-driven models, what is the probability that the requests of these 15 customers can all be met from existing stock?

146. A friend recently planned a camping trip. He has two flashlights, one that required a single 6-V battery and another that used two size-D batteries. He had previously packed two 6-V and four

size-D batteries in his camper. Suppose the probability that any particular battery works is p and that batteries work or fail independently of one another. Our friend wants to take just one flashlight. For what values of p should he take the 6-V flashlight?

147. Binary data are transmitted over a noisy communication channel. The probability that a received binary digit is in error due to channel noise is 0.05. Assume that such errors occur independently within the bit stream.
   (a) What is the probability that the 3rd error occurs on the 50th transmitted bit?
   (b) On average, how many bits will be transmitted correctly *before* the first error?
   (c) Consider a 32-bit "word." What is the probability of exactly 2 errors in this word?
   (d) Consider the next 10,000 bits. What approximating model could we use for $X =$ the number of errors in these 10,000 bits? Give both the name of the model and the value(s) of the parameter(s).

148. A manufacturer of flashlight batteries wishes to control the quality of its product by rejecting any lot in which the proportion of batteries having unacceptable voltage appears to be too high. To this end, out of each large lot (10,000 batteries), 25 will be selected and tested. If at least 5 of these generate an unacceptable voltage, the entire lot will be rejected. What is the probability that a lot will be rejected if
   (a) 5% of the batteries in the lot have unacceptable voltages?
   (b) 10% of the batteries in the lot have unacceptable voltages?
   (c) 20% of the batteries in the lot have unacceptable voltages?
   (d) What would happen to the probabilities in parts (a)–(c) if the critical rejection number were increased from 5 to 6?

149. Of the people passing through an airport metal detector, .5% activate it; let $X =$ the number among a randomly selected group of 500 who activate the detector.
   (a) What is the (approximate) pmf of X?
   (b) Compute $P(X = 5)$.
   (c) Compute $P(X \geq 5)$.

150. An educational consulting firm is trying to decide whether high school students who have never before used a handheld calculator can solve a certain type of problem more easily with a calculator that uses reverse Polish logic or one that does not use this logic. A sample of 25 students is selected and allowed to practice on both calculators. Then each student is asked to work one problem on the reverse Polish calculator and a similar problem on the other. Let $p = P(S)$, where S indicates that a student worked the problem more quickly using reverse Polish logic than without, and let $X =$ number of S's.
   (a) If $p = .5$, what is $P(7 \leq X \leq 18)$?
   (b) If $p = .8$, what is $P(7 \leq X \leq 18)$?
   (c) If the claim that $p = .5$ is to be rejected when either $X \leq 7$ or $X \geq 18$, what is the probability of rejecting the claim when it is actually correct?
   (d) If the decision to reject the claim $p = .5$ is made as in part (c), what is the probability that the claim is not rejected when $p = .6$? When $p = .8$?
   (e) What decision rule would you choose for rejecting the claim $p = .5$ if you wanted the probability in part (c) to be at most .01?

151. Consider a disease whose presence can be identified by carrying out a blood test. Let p denote the probability that a randomly selected individual has the disease. Suppose n individuals are independently selected for testing. One way to proceed is to carry out a separate test on each of the n blood samples. A potentially more economical approach, group testing, was introduced during World War II to identify syphilitic men among army inductees. First, take a part of each

blood sample, combine these specimens, and carry out a single test. If no one has the disease, the result will be negative, and only the one test is required. If at least one individual is diseased, the test on the combined sample will yield a positive result, in which case the n individual tests are then carried out. If $p = .1$ and $n = 3$, what is the expected number of tests using this procedure? What is the expected number when $n = 5$? [The article "Random Multiple-Access Communication and Group Testing" (*IEEE Trans. Commun.*, 1984: 769–774) applied these ideas to a communication system in which the dichotomy was active/idle user rather than diseased/ nondiseased.]

152. Let p_1 denote the probability that any particular code symbol is erroneously transmitted through a communication system. Assume that on different symbols, errors occur independently of one another. Suppose also that with probability p_2 an erroneous symbol is corrected upon receipt. Let X denote the number of correct symbols in a message block consisting of n symbols (after the correction process has ended). What is the probability distribution of X?

153. The purchaser of a power-generating unit requires c consecutive successful start-ups before the unit will be accepted. Assume that the outcomes of individual start-ups are independent of one another. Let p denote the probability that any particular start-up is successful. The random variable of interest is $X =$ the number of start-ups that must be made prior to acceptance. Give the pmf of X for the case $c = 2$. If $p = .9$, what is $P(X \leq 8)$? [*Hint*: For $x \geq 5$, express $p(x)$ "recursively" in terms of the pmf evaluated at the smaller values $x - 3$, $x - 4$, ..., 2.] (This problem was suggested by the article "Evaluation of a Start-Up Demonstration Test," *J. Qual. Tech.*, 1983: 103–106.)

154. A plan for an executive travelers' club has been developed by an airline on the premise that 10% of its current customers would qualify for membership.
    (a) Assuming the validity of this premise, among 25 randomly selected current customers, what is the probability that between 2 and 6 (inclusive) qualify for membership?
    (b) Again assuming the validity of the premise, what are the expected number of customers who qualify and the standard deviation of the number who qualify in a random sample of 100 current customers?
    (c) Let X denote the number in a random sample of 25 current customers who qualify for membership. Consider rejecting the company's premise in favor of the claim that $p > .10$ if $x \geq 7$. What is the probability that the company's premise is rejected when it is actually valid?
    (d) Refer to the decision rule introduced in part (c). What is the probability that the company's premise is not rejected even though $p = .20$ (i.e., 20% qualify)?

155. Forty percent of seeds from maize (modern-day corn) ears carry single spikelets, and the other 60% carry paired spikelets. A seed with single spikelets will produce an ear with single spikelets 29% of the time, whereas a seed with paired spikelets will produce an ear with single spikelets 26% of the time. Consider randomly selecting ten seeds.
    (a) What is the probability that exactly five of these seeds carry a single spikelet and produce an ear with a single spikelet?
    (b) What is the probability that exactly five of the ears produced by these seeds have single spikelets? What is the probability that at most five ears have single spikelets?

156. A trial has just resulted in a hung jury because eight members of the jury were in favor of a guilty verdict and the other four were for acquittal. If the jurors leave the jury room in random order and each of the first four leaving the room is accosted by a reporter in quest of an interview, what is the pmf of $X =$ the number of jurors favoring acquittal among those interviewed? How many of those favoring acquittal do you expect to be interviewed?

157. A reservation service employs five information operators who receive requests for information independently of one another, each according to a Poisson process with rate $\lambda = 2$ per minute.
   (a) What is the probability that during a given 1-min period, the first operator receives no requests?
   (b) What is the probability that during a given 1-min period, exactly four of the five operators receive no requests?
   (c) Write an expression for the probability that during a given 1-min period, all of the operators receive exactly the same number of requests.

158. Grasshoppers are distributed at random in a large field according to a Poisson process with parameter $\lambda = 2$ per square yard. How large should the radius r of a circular sampling region be taken so that the probability of finding at least one grasshopper in the region equals .99?

159. A newsstand has ordered five copies of a certain issue of a photography magazine. Let $X =$ the number of individuals who come in to purchase this magazine. If X has a Poisson distribution with parameter $\mu = 4$, what is the expected number of copies that are sold?

160. Individuals A and B begin to play a sequence of chess games. Let $S = \{$A wins a game$\}$, and suppose that outcomes of successive games are independent with $P(S) = p$ and $P(F) = 1 - p$ (they never draw). They will play until one of them wins ten games. Let $X =$ the number of games played (with possible values 10, 11, ..., 19).
   (a) For $x = 10, 11, \ldots, 19$, obtain an expression for $p(x) = P(X = x)$.
   (b) If a draw is possible, with $p = P(S)$, $q = P(F)$, $1 - p - q = P(\text{draw})$, what are the possible values of X? What is $P(20 \le X)$? [Hint: $P(20 \le X) = 1 - P(X < 20)$.]

161. A test for the presence of a disease has probability .20 of giving a false-positive reading (indicating that an individual has the disease when this is not the case) and probability.10 of giving a false-negative result. Suppose that ten individuals are tested, five of whom have the disease and five of whom do not. Let $X =$ the number of positive readings that result.
   (a) Does X have a binomial distribution? Explain your reasoning.
   (b) What is the probability that exactly three of the ten test results are positive?

162. The generalized negative binomial pmf is given by

$$nb(x; r, p) = k(r, x) \times p^r (1 - p)^x \quad x = 0, 1, 2, \ldots$$

where

$$k(r, x) = \begin{cases} \dfrac{(x + r - 1)(x + r - 2) \ldots (x + r - x)}{x!} & x = 1, 2, \ldots \\ 1 & x = 0 \end{cases}$$

Let X, the number of plants of a certain species found in a particular region, have this distribution with $p = .3$ and $r = 2.5$. What is $P(X = 4)$? What is the probability that at least one plant is found?

163. There are two certified public accountants (CPAs) in a particular office who prepare tax returns for clients. Suppose that for one type of complex tax form, the number of errors made by the first preparer has a Poisson distribution with mean μ_1, the number of errors made by the second preparer has a Poisson distribution with mean μ_2, and that each CPA prepares the same number of forms of this type. Then if one such form is randomly selected, the function

$$p(x; \mu_1, \mu_2) = .5e^{-\mu_1}\frac{\mu_1^x}{x!} + .5e^{-\mu_2}\frac{\mu_2^x}{x!} \quad x = 0, 1, 2, \ldots$$

gives the pmf of $X =$ the number of errors in the selected form.

    (a) Verify that $p(x; \mu_1, \mu_2)$ is a legitimate pmf (≥ 0 and sums to 1).

    (b) What is the expected number of errors on the selected form?

    (c) What is the standard deviation of the number of errors on the selected form?

    (d) How does the pmf change if the first CPA prepares 60% of all such forms and the second prepares 40%?

164. The *mode* of a discrete random variable X with pmf $p(x)$ is that value x^* for which $p(x)$ is largest (the most probable x value).

    (a) Let $X \sim \text{Bin}(n, p)$. By considering the ratio $b(x + 1; n, p)/b(x; n, p)$, show that $b(x; n, p)$ increases with x as long as $x < np - (1 - p)$. Conclude that the mode x^* is the integer satisfying $(n + 1)p - 1 \leq x^* \leq (n + 1)p$.

    (b) Show that if X has a Poisson distribution with parameter μ, the mode is the largest integer less than μ. If μ is an integer, show that both $\mu - 1$ and μ are modes.

165. For a particular insurance policy the number of claims by a policy holder in 5 years is Poisson distributed. If the filing of one claim is four times as likely as the filing of two claims, find the expected number of claims.

166. If X is a hypergeometric rv, show directly from the definition that $E(X) = nM/N$ (consider only the case $n < M$). [*Hint*: Factor nM/N out of the sum for $E(X)$, and show that the terms inside the sum are a match to the pmf $h(y; n - 1, M - 1, N - 1)$, where $y = x - 1$.]

167. Suppose a store sells two different coffee makers of a particular brand, a basic model selling for $30 and a fancy one selling for $50. Let X be the number of people among the next 25 purchasing this brand who choose the fancy one. Then $h(X) = \text{revenue} = 50X + 30(25 - X) = 20X + 750$, a linear function. If the choices are independent and have the same probability, then how is X distributed? Find the mean and standard deviation of $h(X)$. Explain why the choices might not be independent with the same probability.

168. Let X be a discrete rv with possible values 0, 1, 2, ... or some subset of these. The function

$$\psi(s) = E(s^X) = \sum_{x=0}^{\infty} s^x \cdot p(x)$$ is called the **probability generating function (pgf)** of X.

    (a) Suppose X is the number of children born to a family, and $p(0) = .2, p(1) = .5$, and $p(2) = .3$. Determine the pgf of X.

    (b) Determine the pgf when X has a Poisson distribution with parameter μ.

    (c) Show that $\psi(1) = 1$.

    (d) Show that $\psi'(0) = p(1)$. (You'll need to assume that the derivative can be brought inside the summation, which is justified.) What results from taking the second derivative with respect to s and evaluating at $s = 0$? The third derivative? Explain how successive differentiation of $\psi(s)$ and evaluation at $s = 0$ "generates the probabilities in the distribution." Use this to recapture the probabilities of (a) from the pgf. [*Note*: This shows that the pgf contains all the information about the distribution—knowing $\psi(s)$ is equivalent to knowing $p(x)$.]

169. Consider a collection $A_1, \ldots, A_k$ of mutually exclusive and exhaustive events (a partition) and a random variable X whose distribution depends on which of the A_is occurs. (e.g., a commuter might select one of three possible routes from home to work, with X representing commute time.) Let $E(X \mid A_i)$ denote the expected value of X given that event A_i occurs. Then, analogous to the Law of Total Probability, it can be shown that the overall mean of X is given by the weighted average $E(X) = \sum E(X|A_i)P(A_i)$

    (a) The expected duration of a voice call to a particular office telephone number is 3 min, whereas the expected duration of a data call to that same number is 1 min. If 75% of all calls are voice calls, what is the expected duration of the next call?

(b) A bakery sells three different types of chocolate chip cookies. The number of chocolate chips on a type i cookie has a Poisson distribution with mean $\mu_i = i + 1$ ($i = 1, 2, 3$). If 20% of all customers select a cookie of the first type, 50% choose the second type, and 30% opt for the third type, what is the expected number of chocolate chips in the next customer's cookie?

170. Consider a sequence of identical and independent trials, each of which will be a success S or failure F. Let $p = P(S)$ and $q = P(F)$.

(a) Let X = the number of trials necessary to obtain the first S, a geometric rv. Here is an alternative approach to determining $E(X)$. Apply the weighted average formula from the previous exercise with $k = 2$, $A_1 = \{S$ on 1st trial$\}$, and $A_2 = A'$. Show that $E(X) = 1/p$. [*Hint*: Denote $E(X)$ by μ. Given that the first trial is a failure, one trial has been performed and, starting from the 2nd trial, we are still looking for the first S. This implies that $E(X|A') = 1 + \mu$.]

(b) Now let Y = the number of trials necessary to obtain two *consecutive* S's. It is not possible to determine $E(Y)$ directly from the definition of expected value, because there is no formula for the pmf of Y; the complication is the word *consecutive*. Use the weighted average formula to determine $E(Y)$. [*Hint*: Consider the partition with $k = 3$ and $A_1 = \{F\}$, $A_2 = \{SS\}$, $A_3 = \{SF\}$.]