4. Discrete Probability Distributions

4.1. Random Variables and Their Probability Distributions

Most of the experiments we encounter generate outcomes that can be interpreted in terms of real numbers, such as heights of children, numbers of voters favoring various candidates, tensile strength of wires, and numbers of accidents at specified intersections. These numerical outcomes, whose values can change from experiment to experiment, are called *random variables*. We will look at an illustrative example of a random variable before we attempt a more formal definition.

A section of an electrical circuit has two relays, numbered 1 and 2, operating in parallel. The current will flow when a switch is thrown if either one or both of the relays close. The probability that a relay will close properly is 0.8, and the probability is the same for each relay. The relays operate independently, we assume. Let E_i denote the event that relay i closes properly when the switch is thrown. Then $P(E_i) = 0.8$.

When the switch is thrown, a numerical outcome of some interest to the operator of this system is X, the number of relays that close properly. Now, X can take on only three possible values, because the number of relays that close must be 0, 1, or 2. We can find the probabilities associated with these values of X by relating them to the underlying events E_i. Thus, we have

$$\begin{aligned} P(X=0) &= P(\overline{E}_1\overline{E}_2) \\ &= P(\overline{E}_1)P(\overline{E}_2) \\ &= 0.2(0.2) \\ &= 0.04 \end{aligned}.$$

because $X = 0$ means that neither relay closes and the relays operate independently. Similarly,

$$\begin{aligned} P(X=1) &= P(E_1\overline{E}_2 \cup \overline{E}_1E_2) \\ &= P(E_1\overline{E}_2) + P(\overline{E}_1E_2) \\ &= P(E_1)P(\overline{E}_2) + P(\overline{E}_1)P(E_2) \\ &= 0.8(0.2) + 0,2(0.8) \\ &= 0.32 \end{aligned}$$

and

$$\begin{aligned} P(X=2) &= P(E_1E_2) \\ &= P(E_1)P(E_2) \\ &= 0.8(0.8) \\ &= 0.64 \end{aligned}$$

The values of X, along with their probabilities, are more useful for keeping track of the operation of this system than are the underlying events E_i, because the number of properly closing relays is the key to whether the system will work. The current will flow if X is equal to at least 1, and this event has probability

$$\begin{aligned} P(X \geq 1) &= P(X = 1 \ \ or \ \ X = 2) \\ &= P(X = 1) + P(X = 2) \\ &= 0.32 + 0.64 \\ &= 0.96 \end{aligned}$$

Notice that we have mapped the outcomes of an experiment into a set of three meaningful real numbers and have attached a probability to each. Such situations provide the motivation for Definitions 4.1 and 4.2.

> Definition 4.1. A **random variable** is a real-valued function whose domain is a sample space.

Random variables will be denoted by upper-case letters, usually toward the end of the alphabet, such as *X*, *Y*, and *Z*. The actual values that random variables can assume will be denoted by lower-case letters, such as *x*, *y*, and *z*. We can then talk about the "probability that *X* takes on the value *x*," $P(X = x)$, which is denoted by $p(x)$.

In the relay example, the random variable *X* has only three possible values, and it is a relatively simple matter to assign probabilities to these values. Such a random variable is called *discrete*.

> Definition 4.2. A random variable *X* is said to be **discrete** if it can take on only a finite number—or a countably infinite number—of possible values *x*. The **probability function** of *X*, denoted by $p(x)$, assigns probability to each value *x* of *X* so that following conditions are satisfied:
>
> 1. $P(X = x) = p(x) \geq 0$.
> 2. $\sum_x P(X = x) = 1$, where the sum is over all possible values of *x*.

The probability function is sometimes called the **probability mass function** of *X*, to denote the idea that a mass of probability is associated with values for discrete points.

It is often convenient to list the probabilities for a discrete random variable in a table. With *X* defined as the number of closed relays in the problem just discussed, the table is as follows:

x	$p(x)$
0	0.04
1	0.32
2	0.64
Total	1.00

This listing constitutes one way of representing the probability distribution of *X*.

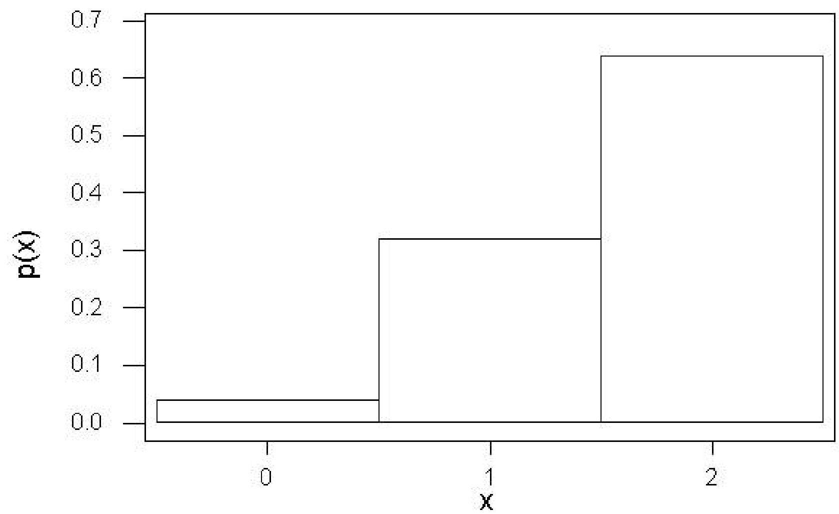
Notice that the probability function $p(x)$ satisfies two properties:

1. $0 \leq p(x) \leq 1$ for any *x*.

2. $\sum_x p(x) = 1$, where the sum is over all possible values of x.

In general, a function is a probability function if and only if the above two conditions are satisfied. Bar graphs are used to display the probability functions for discrete random variables. The probability distribution of the number of closed relays discussed above is shown in Figure 4.1.

Figure 4.1. Graph of a probability mass function



Functional forms for some probability functions that have been useful for modeling real-life data will be given in later sections. We now illustrate another method for arriving at a tabular presentation of a discrete probability distribution.

Example 4.1:

A local video store periodically puts its used movies in a bin and offers to sell them to customers at a reduced price. Twelve copies of a popular movie have just been added to the bin, but three of these are defective. A customer randomly selects two of the copies for gifts. Let X be the number of defective movies the customer purchased. Find the probability function for X and graph the function.

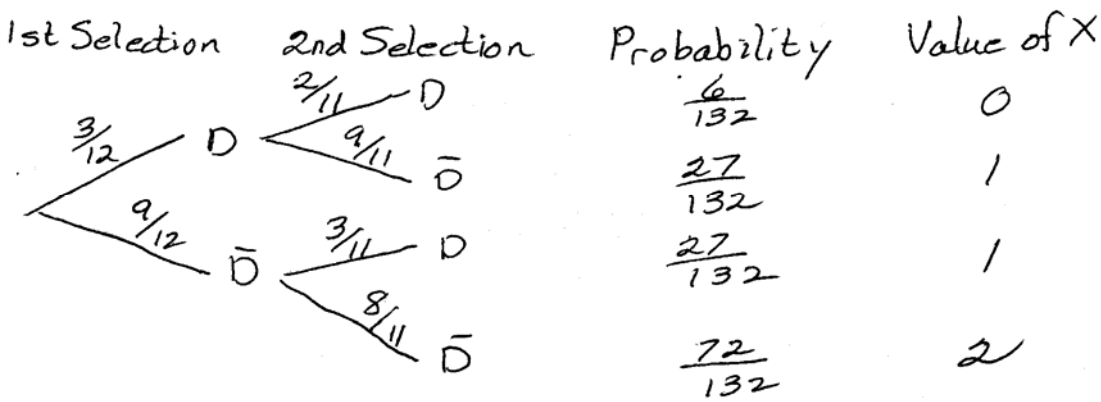
Solution:

The experiment consists of two selections, each of which can result in one of two outcomes. Let D_i denote the event that the ith movie selected is defective; thus, $\overline{D_i}$ denotes the event that it is good. The probability of selecting two good movies ($X = 0$) is

$$P(\overline{D}_1\overline{D}_2) = P(\overline{D}\ on\ 1st)P(\overline{D}\ on\ 2nd \mid \overline{D}\ on\ 1st)$$

The multiplicative law of probability is used, and the probability for the second selection depends on what happened on the first selection. Other possibilities for outcomes will result in other values of X. These outcomes are conveniently listed on the tree in Figure 3.2. The probabilities for the various selections are given on the branches of the tree. Figure 4.2.

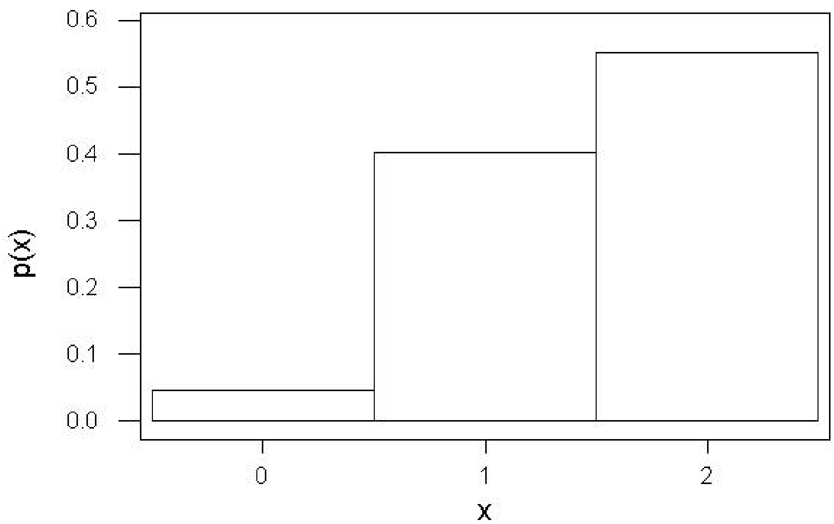
Figure 4.2. Outcomes for Example 4.1



Clearly, X has three possible outcomes, with probabilities as follows:

x	$p(x)$
0	$\frac{72}{132}$
1	$\frac{54}{132}$
2	$\frac{6}{132}$
Total	1.00

The probabilities are graphed in the figure below.



Try to envision this concept extended to more selections from bins of various structures.

We sometimes study the behavior of random variables by looking at the *cumulative* probabilities; that is, for any random variable X, we may look at $P(X \leq b)$ for any real number b. This is, the cumulative probability for X evaluated at b. Thus, we can define a function $F(b)$ as

$$F(b) = P(X \leq b).$$

Definition 4.3. The **distribution function** $F(b)$ for a random variable X is

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

If X is discrete,

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

where $p(x)$ is the probability function.

The distribution function is often called the **cumulative distribution function** (c.d.f).

The random variable X, denoting the number of relays that close properly (as defined at the beginning of this section), has a probability distribution given by

$$P(X = 0) = 0.04$$
$$P(X = 1) = 0.32$$
$$P(X = 2) = 0.64$$

Because positive probability is associated only for $\boldsymbol{x}$ = 0, 1, or 2, the distribution function changes values only at those points. For values of b at least 1, but less than 2, the $P(X \le b) = P(X \le 1)$. For example, we can see that

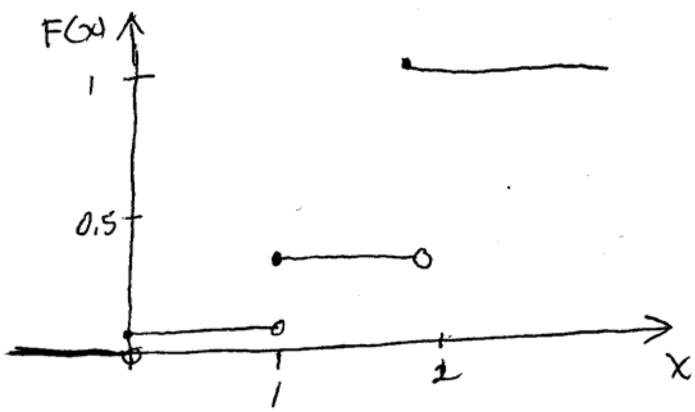
$$P(X \le 1.5) = P(X \le 1.9) = P(X \le 1) = 0.36$$

The distribution function for this random variable then has the form

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.04, & 0 \le x < 1 \\ 0.36, & 1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

The function is graphed in Figure 4.3.

Figure 4.3. Distribution Function



Notice that the distribution function is a step function and is defined for all real numbers. This is true for all discrete random variables. The distribution function is discontinuous at points of positive probability. Because the outcomes 0, 1, and 2 have positive probability associated with them, the distribution function is discontinuous at those points. The change in the value of the function at a point (the height of the step) is the probability associated with that value x. Since the outcome of 2 is the most probable ($p(2) = 0.64$), the height of the "step" at this point is the largest. Although the function has points of discontinuity, it is right-hand continuous at all points. To see this, consider $X = 1$. As we approach 1 from the left, we have $\lim_{h \to 0^+} F(1+h) = 0.36 = F(1)$; that is, the distribution function F is right-hand continuous. However, if we approach 1 from the left, we find $\lim_{h \to 0^-} F(1+h) = 0.04 \ne 0.36 = F(1)$, giving us that F is not left-hand continuous. Because a function must be both left- and right-hand continuous to be continuous, F is not continuous at $X = 1$.

In general, a distribution function is defined for the whole real line. Every distribution function must satisfy four properties; similarly, any function satisfying the following four properties is a distribution function.

1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to \infty} F(x) = 1$
3. The distribution function is a non-decreasing function; that is, if $a < b$, $F(a) \leq F(b)$. The distribution function can remain constant, but it cannot decrease, as we increase from a to b.
4. The distribution function is right-hand continuous; that is, $\lim_{h \to 0^+} F(x+h) = F(x)$

We have already seen that, given a probability mass function, we can determine the distribution function. For any distribution function, we can also determine the probability function.

Example 4.2:

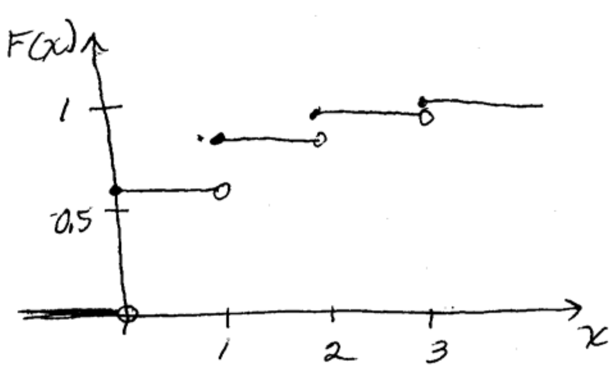
A large university uses some of the student fees to offer free use of its Health Center to all students. Let X be the number of times that a randomly selected student visits the Health Center during a semester. Based on historical data, the distribution function of X is given below.

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.6, & 0 \leq x < 1 \\ 0.8, & 1 \leq x < 2 \\ 0.95, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

For the function above,
1. Graph F.
2. Verify that F is a distribution function.
3. Find the probability function associated with F.

Solution:

1.



2. To verify F is a distribution function, we must confirm that the function satisfies the four conditions of a distribution function.

(1) Because F is zero for all values x less than 0, $\lim_{x \to -\infty} F(x) = 0$.

(2) Similarly F is one for all values of x that are 3 or greater; therefore, $\lim_{x \to +\infty} F(x) = 1$.

(3) The function F is non-decreasing. There are many points for which it is not increasing, but as x increases, $F(x)$ either remains constant or increases.

(4) The function is discontinuous at four points: 0, 1, 2, and 3. At each of these points, F is right-hand continuous. As an example, for $X = 2$, $\lim_{h \to 0^+} F(2+h) = 0.95 = F(2)$.

Because F satisfies the four conditions, it is a distribution function.

3. The points of positive probability occur at the points of discontinuity: 0, 1, 2, and 3. Further, the probability is the height of the "jump" at that point. This gives us the following probabilities.

x	$p(x)$
0	$0.6 - 0 = 0.6$
1	$0.8 - 0.6 = 0.2$
2	$0.95 - 0.8 = 0.15$
3	$1 - 0.95 = 0.5$

Exercises

4.1. Circuit boards from two assembly lines set up to produce identical boards are mixed in one storage tray. As inspectors examine the boards, they find that it is difficult to determine whether a board comes from line A or line B. A probabilistic assessment of this question is often helpful. Suppose that the storage tray contains ten circuit boards, of which six came from line A and four from line B. An inspector selects two of these identical-looking boards for inspection. He is interested in X, the number of inspected boards from line A.

a. Find the probability function for X.
b. Graph the probability function of X.
c. Find the distribution function of X.
d. Graph the distribution function of X.

4.2. Among twelve applicants for an open position, seven are women and five are men. Suppose that three applicants are randomly selected from the applicant pool for final interviews. Let X be the number of female applicants among the final three.

a. Find the probability function for X.
b. Graph the probability function of X.
c. Find the distribution function of X.
d. Graph the distribution function of X.

4.3. The median annual income for heads of households in a certain city is \$44,000. Four such heads of household are randomly selected for inclusion in an opinion poll. Let X be the number (out of the four) who have annual incomes below \$44,000.
a. Find the probability distribution of X.
b. Graph the probability distribution of X.
c. Find the distribution function of X.
d. Graph the distribution function of X.
e. Is it unusual to see all four below \$44,000 in this type of poll? (What is the probability of this event?)

4.4. At a miniature golf course, players record the strokes required to make each hole. If the ball is not in the hole after five strokes, the player is to pick up the ball and record six strokes. The owner is concerned about the flow of players at hole 7. (She thinks that players tend to get backed up at that hole.). She has determined that the distribution function of X, the number of strokes a player takes to complete hole 7 to be

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.05, & 1 \le x < 2 \\ 0.15, & 2 \le x < 3 \\ 0.35, & 3 \le x < 4 \\ 0.65, & 4 \le x < 5 \\ 0.85, & 5 \le x < 6 \\ 1, & x \ge 6 \end{cases}$$

a. Graph the distribution function of X.
b. Find the probability function of X.
c. Graph the probability function of X.
d. Based on (a) through (c), are the owner's concerns substantiated?

4.5. In 2005, Derrek Lee led the National Baseball League with a 0.335 batting average, meaning that he got a hit on 33.5% of his official times at bat. In a typical game, he had three official at bats.
a. Find the probability distribution for X, the number of hits Boggs got in a typical game.
b. What assumptions are involved in the answer? Are the assumptions reasonable?
c. Is it unusual for a good hitter to go 0 for 3 in one game?

4.6. A commercial building has three entrances, numbered I, II, and III. Four people enter the building at 9:00 a.m. Let X denote the number who select entrance I. Assuming that the people choose entrances independently and at random, find the probability distribution for X. Were any additional assumptions necessary for your answer?

4.7. In 2002, 33.9% of all fires were structure fires. Of these, 78% of these were residential fires. The causes of structure fire and the numbers of fires during 2002 for each cause are displayed in the table below. Suppose that four independent structure fires are reported in one day, and let X denote the number, out of the four, that are caused by cooking.

Cause of Fire	Number of Fires
Cooking	29,706
Chimney Fires	8,638
Incinerator	284
Fuel Burner	3,226
Commercial Compactor	246
Trash/Rubbish	9,906

a. Find the probability distribution for X, in tabular form.
b. Find the probability that at least one of the four fires was caused by cooking.

4.8. Observers have noticed that the distribution function of X, the number of commercial vehicles that cross a certain toll bridge during a minute is as follows:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.20, & 0 \le x < 1 \\ 0.50, & 1 \le x < 2 \\ 0.85, & 2 \le x < 4 \\ 1, & x \ge 4 \end{cases}$$

a. Graph the distribution function of X.
b. Find the probability function of X.
c. Graph the probability function of X.

4.9. Of the people who enter a blood bank to donate blood, 1 in 3 have type O^+ blood, and 1 in 20 have type O^- blood. For the next three people entering the blood bank, let X denote the number with O^+ blood, and let Y denote the number with O^- blood. Assume the independence among the people with respect to blood type.
a. Find the probability distribution for X and Y.
b. Find the probability distribution of $X + Y$, the number of people with type O blood.

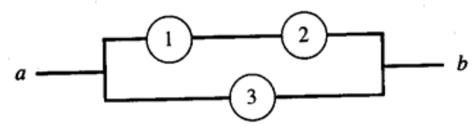
4.10. Daily sales records for a car dealership show that it will sell 0, 1, 2, or 3 cars, with probabilities as listed:

Number of Sales	0	1	2	3
Probability	0.5	0.3	0.15	0.05

a. Find the probability distribution for *X*, the number of sales in a two-day period, assuming the sales are independent from day to day.
b. Find the probability that at least one sale is made in the next two days.

4.11. Four microchips are to be placed in a computer. Two of the four chips are randomly selected for inspection before the computer is assembled. Let *X* denote the number of defective chips found among the two inspected. Find the probability distribution for *X* for the following events.
a. Two of the microchips were defective.
b. One of the four microchips was defective.
c. None of the microchips was defective.

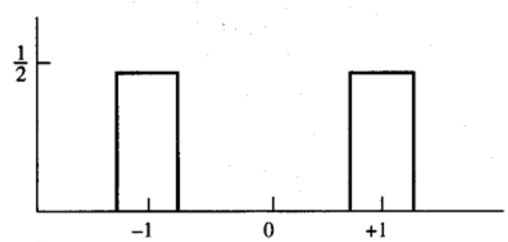
4.12. When turned on, each of the three switches in the accompanying diagram works properly with probability 0.9. If a switch is working properly, current can flow through it when it is turned on. Find the probability distribution for *Y*, the number of closed paths from *a* to *b*, when all three switches are turned on.



4.2 Expected Values of Random Variables

Because a probability can be thought of as the long-run relative frequency of occurrence for an event, a probability distribution can be interpreted as showing the long-run relative frequency of occurrences for numerical outcomes associated with an experiment. Suppose, for example, that you and a friend are matching balanced coins. Each of you flips a coin. If the upper faces match, you win $1.00; if they do not match, you lose $1.00 (your friend wins $1.00). The probability of a match is 0.5 and, in the long run, you should win about half of the time. Thus, a relative frequency distribution of your winnings should look like the one shown in Figure 4.4. The -1 under the left most bar indicates a loss of $1.00 by you.

Figure 4.4. Relative frequency of winnings



On average, how much will you win per game over the long run? If Figure 4.4 presents a correct display of your winnings, you win -1 (lose a dollar) half of the time and +1 half of the time, for an average of

$$(-1)\left(\frac{1}{2}\right)+(1)\left(\frac{1}{2}\right)=0$$

This average is sometimes called your expected winnings per game, or the expected value of your winnings. (A game that has an expected value of winnings of 0 is called a fair game.) The general definition of expected value is given in Definition 4.4.

Definition 4.4. The **expected value** of a discrete random variable X with probability distribution $p(x)$ is given by

$$E(X)=\sum_{x} xp(x)$$

(The sum is over all values of x for which $p(x) > 0$.)

We sometimes use the notation

$$E(X) = \mu$$

for this equivalence.

Note: We assume absolute convergence when the range of X is countable; we talk about an expectation only when it is assumed to exist..

.

Now payday has arrived, and you and your friend up the stakes to \$10 per game of matching coins. You now win -10 or +10 with equal probability. Your expected winnings per game is

$$(-10)\left(\frac{1}{2}\right)+(10)\left(\frac{1}{2}\right)=0$$

and the game is still fair. The new stakes can be thought of as a function of the old in the sense that, if X represents your winnings per game when you were playing for \$1.00, then $10X$ represents your winnings per game when you play for \$10.00. Such functions of random variables arise often. The extension of the definition of expected value to cover these cases is given in Theorem 4.1.

Theorem 4.1. If X is a discrete random variable with probability distribution $p(x)$ and if $g(x)$ is any real-valued function of X, then

$$E(g(X))=\sum_{x} g(x)p(x)$$

(The proof of this theorem will not be given.)

You and your friend decide to complicate the payoff rules to the coin-matching game by agreeing to let you win \$1 if the match is tails and \$2 if the match is heads. You still lose $1 if the coins do not match. Quickly you see that this is not a fair game, because your expected winnings are

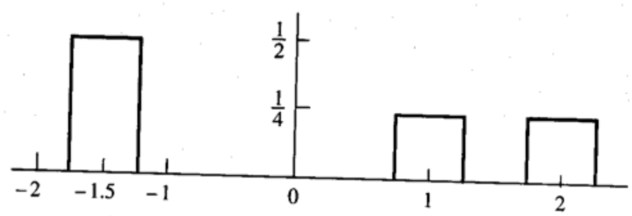
$$(-1)\left(\frac{1}{2}\right)+(1)\left(\frac{1}{4}\right)+(2)\left(\frac{1}{4}\right)=0.25$$

You compensate for this by agreeing to pay your friend $1.50 if the coins do not match. Then, your expected winnings per game are

$$(-1.5)\left(\frac{1}{2}\right)+(1)\left(\frac{1}{4}\right)+(2)\left(\frac{1}{4}\right)=0$$

and the game is again fair. What is the difference between this game and the original one, in which all payoffs were $1? The difference certainly cannot be explained by the expected value, since both games are fair. You can win more, but also lose more, with the new payoffs, and the difference between the two games can be explained to some extent by the increased variability of your winnings across many games. This increased variability can be seen in Figure 4.5, which displays the relative frequency for your winnings in the new game; the winnings are more spread out than they were in Figure 4.4. Formally, variation is often measured by the variance and by a related quantity called the standard deviation.

Figure 4.5. Relative frequency of winnings



Definition 4.5. The **variance** of a random variable X with expected value μ is given by

$$V(X)=E\left[(X-\mu)^2\right].$$

We sometimes use the notation

$$E\left[(X-\mu)^2\right]=\sigma^2$$

for this equivalence.

Notice that the variance can be thought of as the average squared distance between values of X and the expected value μ. Thus, the units associated with σ^2 are the square of the units of measurement for X. The smallest value that σ^2 can assume is zero. The variance is zero when all the probability is concentrated at a single point, that is, when X takes on a constant value with probability 1. The variance becomes larger as the points with positive probability spread out more.

The standard deviation is a measure of variation that maintains the original units of measure, as opposed to the squared units associated with the variance.

Definition 4.6. The **standard deviation** of a random variable X is the square root of the variance and is given by

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

For the game represented in Figure 4.4, the variance of you winnings (with μ = 0) is

$$\sigma^2 = E\left[(X-\mu)^2\right]$$
$$= (-1)^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{2}\right) = 1$$

It follows that σ = 1, as well. For the game represented in Figure 4.5, the variance of your winnings is

$$\sigma^2 = (-1.5)^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{4}\right) + 2^2\left(\frac{1}{4}\right)$$
$$= 2.375$$

and the standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.375} = 1.54$$

Which game would you rather play?

The standard deviation can be thought of as the size of a “typical” deviation between an observed outcome and the expected value. For the situation displayed in Figure 4.4, each outcome (-1 or +1) deviates by precisely one standard deviation from the expected value. For the situation described in Figure 4.5, the positive values average 1.5 units from the expected value of 0 (as do the negative values), and so 1.5 units is approximately one standard deviation here.

The mean and the standard deviation often yield a useful summary of the probability distribution for a random variable that can assume many values. An illustration is provided by the age distribution of the U.S. population for 2000 and 2100 (projected, as shown in Table 4.1).

Age is actually a continuous measurement, but since it is reported in categories, we can treat it as a discrete random variable for purposes of approximating its key function. To move from continuous age intervals to discrete age classes, we assign each interval the value of its midpoint (rounded). Thus, the data in Table 4.1 are interpreted as showing that 6.9% of the 2000 population were around 3 years of age and that 11.6% of the 2100 population is anticipated to be around 45 years of age. (The open intervals at the upper end were stopped at 100 for convenience.)

Table 4.1. Age Distribution in 2000 and 2100 (Projected)

Age Interval	Age Midpoint	2000	2100
Under 5	3	6.9	6.3
5—9	8	7.3	6.2
10—19	15	14.4	12.8
20—29	25	13.3	12.3
30—39	35	15.5	12.0
40—49	45	15.3	11.6
50—59	55	10.8	10.8
60—69	65	7.3	9.8
70—79	75	5.9	8.3
80 and over	90	3.3	9.9

*Source: U.S. Census Bureau

Interpreting the percentages as probabilities, we see that the mean age for 2000 is approximated by

$$\begin{aligned}\mu &= \sum_x xp(x) \\ &= 3(0.069) + 8(0.073) + 15(0.144) + ... + 90(0.033) \\ &= 36.6\end{aligned}$$

(How does this compare with the median age for 2000, as approximated from Table 4.1.) For 2100, the mean age is approximated by

$$\begin{aligned}\mu &= \sum_x xp(x) \\ &= 3(0.062) + 8(0.062) + 15(0.128) + ... + 90(0.099) \\ &= 42.5\end{aligned}$$

Over the projected period, the mean age increases rather markedly (as does the median age).

The variations in the two age distributions can be approximated by the standard deviations. For 2000, this is

$$\begin{aligned}\sigma &= \sqrt{\sum_x (x-\mu)^2 p(x)} \\ &= (3-36.6)^2(0.069) + (8-36.6)^2(0.073) + (15-36.6)^2(0.144) + ... + (90-36.6)^2(0.033) \\ &= 22.6\end{aligned}$$

A similar calculation for the 2100 data yields σ = 26.3. These results are summarized in Table 4.2.

Table 4.2. Age Distribution of U.S. Population Summary

	2000	2100
Mean	36.6	42.5
Standard Deviation	22.6	26.3

Not only is the population getting older, on average, but its variability is increasing. What are some of the implications of these trends?

We now provide other examples and extensions of these basic results.

Example 4.3:

A department supervisor is considering purchasing a photocopy machine. One consideration is how often the machine will need repairs. Let X denote the number of repairs during a year. Based on past performance, the distribution of X is shown below:

Number of repairs, x	0	1	2	3
$p(x)$	0.2	0.3	0.4	0.1

1. What is the expected number of repairs during a year?
2. What is the variance of the number of repairs during a year?

Solution:

1. From Definition 4.4, we see that

$$E(X) = \sum_x xp(x)$$
$$= 0(0.2) + 1(0.3) + 2(0.4) + 3(0.1)$$
$$= 1.4$$

The photocopier will need to be repaired an average of 1.4 times per year.

2. From Definition 4.5, we see that

$$V(X) = \sum_x (x-\mu)^2 p(x)$$
$$= (0-1.4)^2(0.2) + (1-1.4)^2(0.3) + (2-1.4)^2(0.4) + (3-1.4)^2(0.1)$$
$$= 0.84$$

Our work in manipulating expected values can be greatly facilitated by making use of the two results of Theorem 4.2. Often, $g(X)$ is a linear function. When that is the case, the calculations of expected value and variance are especially simple.

Theorem 4.2. For any random variable X and constants a and b,

1. $E(aX + b) = aE(X) + b$
2. $F(aX + b) = aE(X) + b$

Proof:
We sketch a proof of this theorem for a discrete random variable X having a probability distribution given by p(x). By Theorem 4.1,

$$\begin{aligned} E(aX+b) &= \sum_x (ax+b)p(x) \\ &= \sum_x [(ax)p(x) + bp(x)] \\ &= \sum_x axp(x) + \sum_x bp(x) \\ &= a\sum_x xp(x) + b\sum_x p(x) \\ &= aE(X) + b \end{aligned}$$

Notice that $\sum_x p(x)$ must equal unity. Also, by Definition 4.5,

$$\begin{aligned} V(aX+b) &= E[(aX+b) - E(aX+b)]^2 \\ &= E[aX + b - (aE(X)+b)]^2 \\ &= E[aX - aE(X)]^2 \\ &= E\left[a^2(x - E(X))^2\right] \\ &= a^2 E\left[(X - E(X))^2\right] \\ &= a^2 V(X) \end{aligned}$$

An important special case of Theorem 4.2 involves establishing a "standardized variable. If X has mean μ and standard deviation σ, then the "standardized" form of X is given by

$$Y = \frac{X - \mu}{\sigma}$$

Employing Theorem 4.2, one can show that $E(Y) = 0$ and $V(Y^2) = 1$. This idea will be used often in later chapters.

We illustrate the use of these results in the following example.

Example 4.4:

The department supervisor in Example 4.3 wants to consider the cost of maintenance before purchasing the photocopy machine. The cost of maintenance consists of the expense of a service agreement and the cost of repairs. The service agreement can be purchased for \$200. With the agreement, the cost of each repair is \$50. Find the mean and variance of the annual costs of repair for the photocopier

Solution:

Recall that the X of Example 4.3 is the annual number of repairs. The annual cost of the maintenance contract is $50X + 200$. By Theorem 4.2, we have

$$
\begin{aligned}
E(50X+200) &= 50E(X)+200 \\
&= 50(1.4)+200 \\
&= 270
\end{aligned}
$$

Thus, the manager could anticipate the average annual cost of maintenance of the photocopier to be \$270.

Also, by Theorem 4.2,

$$
\begin{aligned}
V(50X+200) &= 50^2 V(X) \\
&= 50^2(0.84) \\
&= 2100
\end{aligned}
$$

We will make use of this value in a later example.

Determining the variance by Definition 4.5 is not computationally efficient. Theorem 4.2 leads us to a more efficient formula for computing the variance as given in Theorem 4.3.

Theorem 4.3. If X is a random variable with mean μ, then

$$V(X) = E\left(X^2\right) - \mu^2$$

Proof:
Starting with the definition of variance, we have

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

Example 4.5:

Use the result of Theorem 4.3 to compute the variance of *X* as given in Example 4.3.

Solution:

In Example 4.3, *X* had a probability distribution given by

x	0	1	2	3
$p(x)$	0.2	0.3	0.4	0.1

and we found that $E(X) = 1.4$. Now,

$$E(X^2) = \sum_x x^2 p(x)$$

$$= 0^2(0.2) + 1^2(0.3) + 2^2(0.4) + 3^2(0.1)$$

$$= 2.8$$

By Theorem 4.3,

$$V(X) = E(X^2) - \mu^2$$

$$= 2.8 - (1.4)^2$$

$$= 0.84$$

We have computed means and variances for a number of probability distributions and noted that these two quantities give us some useful information on the center and spread of the probability mass. Now suppose that we know only the mean and the variance for a probability distribution. Can we say anything specific about probabilities for certain intervals about the mean? The answer is "yes," and a useful result of the relationship among mean, standard deviation, and relative frequency will now be discussed.

The inequality in the statement of the theorem is equivalent to

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

To interpret this result, let $k = 2$, for example. Then the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least $1 - 1/k^2 = 1 - ¼ = 3/4$ of the probability mass for the random variable. We consider more specific illustrations in the following two examples.

Theorem 4.4: **Tchebysheff's Theorem**. Let X be a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X-\mu|<k\sigma)\geq 1-\frac{1}{k^2}$$

Proof:
We begin with the definition of $V(X)$ and then make substitutions in the sum defining this quantity. Now,

$$\begin{aligned} V(X) &= \sigma^2 \\ &= \sum_{-\infty}^{\infty}(x-\mu)^2 p(x) \\ &= \sum_{-\infty}^{\mu-k\sigma}(x-\mu)^2 p(x)+\sum_{\mu-k\sigma}^{\mu+k\sigma}(x-\mu)^2 p(x)+\sum_{\mu+k\sigma}^{\infty}(x-\mu)^2 p(x) \end{aligned}$$

(The first sum stops at the largest value of x smaller than $\mu - k\sigma$, and the third sum begins at the smallest value of x larger than $\mu + k\sigma$; the middle sum collects the remaining terms.) Observe that the middle sum is never negative; and for both of the outside sums,

$$(x-\mu)^2 \geq k^2\sigma^2$$

Eliminating the middle sum and substituting for $(x-\mu)^2$ in the other two, we get

$$\sigma^2 \geq \sum_{-\infty}^{\mu-k\sigma} k^2\sigma^2 p(x)+\sum_{\mu+k\sigma}^{\infty} k^2\sigma^2 p(x)$$

or

$$\sigma^2 \geq k^2\sigma^2\left[\sum_{-\infty}^{\mu-k\sigma} p(x)+\sum_{\mu+k\sigma}^{\infty} p(x)\right]$$

or

$$\sigma^2 \geq k^2\sigma^2 P\big(|X-\mu|\geq k\sigma\big).$$

It follows that

$$P\big(|X-\mu|\geq k\sigma\big)\leq \frac{1}{k^2}$$

or

$$P\big(|X-\mu|<k\sigma\big)\geq 1-\frac{1}{k^2}$$

Example 4.6:

The daily production of electric motors at a certain factory averaged 120, with a standard deviation of 10.
1. What can be said about the fraction of days on which the production level falls between 100 and 140?
2. Find the shortest interval certain to contain at least 90% of the daily production levels.

Solution:

1. The interval from 100 to 140 represents μ - 2σ to μ + 2σ, with μ = 120 and σ = 10. Thus, $k = 2$ and

$$1-\frac{1}{k^2}=1-\frac{1}{4}=\frac{3}{4}$$

At least 75% of all days, therefore, will have a total production value that falls in this interval. (This percentage could be closer to 95% if the daily production figures show a mound-shaped, symmetric relative frequency distribution.)

2. To find k, we must set $\left(1-1/k^2\right)$ equal to 0.9 and solve for k; that is,

$$1-\frac{1}{k^2}=0.9$$

$$\frac{1}{k^2}=0.1$$

$$k^2=10$$

$$k=\sqrt{10}$$

$$=3.16$$

The interval

$$\mu-3.16\sigma \text{ to } \mu+3.16\sigma$$

or

$$120-3.16(10) \text{ to } 120+3.16(10)$$

or

$$88.4 \text{ to } 151.6$$

will then contain at least 90% of the daily production levels.

Example 4.7:

The annual cost of maintenance for a certain photocopy machine has a mean of \$270 and a variance of \$2100 (see Example 4.5). The manager wants to budget enough for maintenance that he is unlikely to go over the budgeted amount. He is considering budgeting \$400 for maintenance. How often will the maintenance cost exceed this amount?

Solution:

First, we must find the distance between the mean and 400, in terms of the standard deviation of the distribution of costs. We have

$$\frac{400-\mu}{\sqrt{\sigma^2}} = \frac{400-270}{\sqrt{2100}} = \frac{130}{45.8} = 2.84$$

Thus, 400 is 2.84 standard deviations aboe the mean. Letting $k = 2.84$ in Theorem 4.4, we can find the interval

$$\mu - 2.84\sigma \text{ to } \mu + 2.84\sigma$$

or

$$270 - 2.84(45.8) \text{ to } 270 + 2.84(45.8)$$

or

$$140 \text{ to } 400$$

must contain at least

$$1-\frac{1}{k^2} = 1-\frac{1}{(2.84)^2} = 1-0.12 = 0.88$$

of the probability. Because the annual cost is \$200 plus \$50 for each repair, the annual cost cannot be less than \$200. Thus, at most 0.12 of the probability mass can exceed \$400; that is, the cost cannot exceed \$400 more than 12% of the time.

Example 4.8:

Suppose the random variable X has the probability mass function given in the table below.

x	-1	0	1
$p(x)$	1/8	3/4	1/8

Evaluate Tchebysheff's inequality for $k = 1$.

Solution:

First, we find the mean of X

$$\mu = \sum_x xp(x) = (-1)(1/8) + 0(3/4) + 1(1/8) = 0$$

Then

$$E(X^2) = \sum_x x^2 p(x) = (-1)^2(1/8) + 0^2(3/4) + 1^2(1/8) = 1/4$$

and

$$\sigma^2 = E(X^2) - \mu^2 = \tfrac{1}{4} - 0 = \tfrac{1}{4}$$

Thus, the standard deviation of X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\tfrac{1}{4}} = \tfrac{1}{2}$$

Now, for X, the probability X is within 2σ of μ is

$$\begin{aligned} P(|X-\mu|<2\sigma) &= P(|X-\mu|<2(1/2)) \\ &= P(|X-\mu|<1) \\ &= P(X=0) \\ &= 3/4 \end{aligned}$$

By Tchebysheff's theorem, the probability any random variable X is within 2σ of μ is

$$P(|X-\mu|<2\sigma) \geq 1-\frac{1}{k^2} = 1-\frac{1}{2^2} = \frac{3}{4}$$

Therefore, for this particular random variable X, equality holds in Tchebysheff's theorem. Thus, one cannot improve on the bounds of the theorem.

Exercises

4.13. You are to pay $1.99 to play a game that consists of drawing one ticket at random from a box of unnumbered tickets. You win the amount (in dollars) of the number on the ticket you draw. The following two boxes of numbered tickets are available.

I. $\boxed{0, 1, 2}$ II. $\boxed{0, 0, 0, 1, 4}$

a. Find the expected value and variance of your net gain per play with box I.
b. Repeat part (a) for box II.
c. Given that you have decided to play, which box would you choose, and why?

4.14. The size distribution of U.S. families is shown in the table below.

Number of Persons	Percentage
1	25.7%
2	32.2
3	16.9
4	15.0
5	6.9
6	2.2
7 or more	1.2

a. Calculate the mean and the standard deviation of family size. Are these exact values or approximations?
b. How does the mean family size compare to the median family size?

4.15. The table below shows the estimated number of AIDS cases in the United States by age group.

Numbers of AIDS Cases in the U.S. during 2004

Age	Number of Cases
Under 14	108
15 to 19	326
20 to 24	1,788
25 to 29	3,576
30 to 34	4,786
35 to 39	8,031
40 to 44	8,747
45 to 49	6,245
50 to 54	3,932
55 to 59	2,079
60 to 64	996
65 or older	901
Total	41,515

Source: U.S. Centers for Disease Control

Let X denote the age of a person with AIDS.
a. Using the mid-point of the interval to represent the age of all individuals in that age category, find the approximate probability distribution for X
b. Approximate the mean and the standard deviation of this age distribution.
c. How does the mean age compare to the approximate median age?

4.16. How old are our drivers? The accompanying table gives the age distribution of licensed drivers in the United States. Describe this age distribution in terms of median, mean, and standard deviation.

Licensed U.S. Drivers in 2004

Age	Number (in millions)
19 and under	9.3
20-24	16.9
25-29	17.4
30-34	18.7
35-39	19.4
40-44	21.3
45-49	20.7
50-54	18.4
55-59	15.8
60-64	11.9
65-69	9.0
70-74	7.4
75-79	6.1
80-84	4.1
85 and over	2.5
Total	198.9

Source: U.S. Department of Transportation

4.17. Who commits the crimes in the United States? Although this is a very complex question, one way to address it is to look at the age distribution of those who commit violent crimes. This is presented in the table below. Describe the distribution in terms of median, mean, and standard deviation.

Age	Percent of Violent Crimes
14 and Under	5.1
15-19	19.7
20-24	20.2
25-29	13.6
30-34	12.0
35-39	10.8
40-44	8.7
45-49	5.1
50-54	2.5
55-59	1.2
60-64	0.6
65 and Older	0.5

4.18. A fisherman is restricted to catching at most 2 red grouper per day when fishing in the Gulf of Mexico. A field agent for the wildlife commission often inspects the day's catch for boats as they come to shore near his base. He has found the number of grouper caught has the following distribution.

Number of Grouper	0	1	2
Probability	0.2	0.7	0.1

Assuming that these records are representative of red grouper daily catches in the Gulf, find the expected value, the variance, and the standard deviation for the individual daily catch of red grouper.

4.19. Approximately 10% of the glass bottles coming off a production line have serious defects in the glass. Two bottles are randomly selected for inspection. Find the expected value and the variance of the number of inspected bottles with serious defects.

4.20. Two construction contracts are to be randomly assigned to one or more of three firms—I, II, and III. A firm may receive more than one contract. Each contract has a potential profit of $90,000.

a. Find the expected potential profit for firm I.

b. Find the expected potential profit for firms I and II together.

4.21. Two balanced coins are tossed. What are the expected value and the variance of the number of heads observed?

4.22. In a promotional effort, new customers are encouraged to enter an on-line sweepstakes. To play, the new customer picks 9 numbers between 1 and 50, inclusive. At the end of the promotional period, 9 numbers from 1 to 50, inclusive, are drawn without replacement from a hopper. If the customer's 9 numbers match all of those drawn (without concern for order), the customer wins $5,000,000.
a. What is the probability that the new customer wins the $5,000,000?
b. What is the expected value and variance of the winnings?
c. If the new customer had to mail in the picked numbers, assuming that the cost of postage and handling is $0.50, what is the expected value and variance of the winnings?

4.23. The number of equipment breakdowns in a manufacturing plant is closely monitored by the supervisor of operations, since it is critical to the production process. The number averages 5 per week, with a standard deviation of 0.8 per week.
a. Find an interval that includes at least 90% of the weekly figures for number of breakdowns.
b. The supervisor promises that the number of breakdowns will rarely exceed 8 in a one-week period. Is the director safe in making this claim? Why?

4.24. Keeping an adequate supply of spare parts on hand is an important function of the parts department of a large electronics firm. The monthly demand for 100-gigabyte hard drives for personal computers was studied for some months and found to average 28 with a standard deviation of 4. How many hard drives should be stocked at the beginning of each month to ensure that the demand will exceed the supply with a probability of less than 0.10?

4.25. An important feature of golf cart batteries is the number of minutes they will perform before needing to be recharged. A certain manufacturer advertises batteries that will run, under a 75-amp discharge test, for an average of 125 minutes, with a standard deviation of 5 minutes.
a. Find an interval that contains at least 90% of the performance periods for batteries of this type.
b. Would you expect many batteries to die out in less than 100 minutes? Why?

4.26. Costs of equipment maintenance are an important part of a firm's budget. Each visit by a field representative to check out a malfunction in a certain machine used in a manufacturing process is $65, and the parts cost, on average, about $125 to correct each malfunction. In this large plant, the expected number of these machine malfunctions is approximately 5 per month, and the standard deviation of the number of malfunctions is 2.
a. Find the expected value and standard deviation of the monthly cost of visits by the field representative.
b. How much should the firm budget per month to ensure that the costs of these visits are covered at least 75% of the time?

At this point, it may seem that every problem has its own unique probability distribution, and that we must start from basics to construct such a distribution each time a new problem comes up. Fortunately, this is not the case. Certain basic probability distributions can be developed as models for a large number of practical problems. In the remainder of this chapter, we shall consider some fundamental discrete distributions, looking at the theoretical assumptions that underlie these distributions as well as at the means, variances, and applications of the distributions.

4.3 The Bernoulli Distribution

Numerous experiments have two possible outcomes. If an item is selected from the assembly line and inspected, it will be found to be either defective or not defective. A piece of fruit is either damaged or not damaged. A cow is either pregnant or not pregnant. A child will be either female or male. Such experiments are called *Bernoulli trials* after the Swiss mathematician Jacob Bernoulli.

For simplicity, suppose one outcome of a Bernoulli trial is identified to be a success and the other a failure. Define the random variable X as follows:

$$\begin{aligned} X &= 1, \text{ if the outcome of the trial is a success} \\ &= 0, \text{ if the outcome of the trial is a failure} \end{aligned}$$

If the probability of observing a success is p, the probability of observing failure is $1 - p$. The probability distribution of X, then, is given by

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

where $p(x)$ denotes the probability that $X = x$. Such a random variable is said to have a Bernoulli distribution or to represent the outcome of a single Bernoulli trial. A general formula for $p(x)$ identifies a family of distributions indexed by certain constants called parameters. For the Bernoulli distribution, the probability of success, p, is the only parameter.

Suppose that we repeatedly observe the outcomes of random experiments of this type, recording a value of X for each outcome. What average of X should we expect to see? By Definition 4.4, the expected value of X is given by

$$\begin{aligned} E(X) &= \sum_x xp(x) \\ &= 0p(0) + 1(p(1) \\ &= 0(1-p) + 1(p) = p \end{aligned}$$

Thus, if we inspect a single item from an assembly line and 10% of the items are defective, we should expect to observe an average of 0.1 defective items per item inspected. (In other words, we should expect to see one defective item for every ten items inspected.)

For the Bernoulli random variable X, the variance (see Theorem 4.3) is

$$
\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \sum_x x^2 p(x) - p^2 \\
&= 0^2(1-p) + 1^2(p) - p^2 \\
&= p - p^2 = p(1-p)
\end{aligned}
$$

Seldom is one interested in observing only one outcome of a Bernoulli trial. However, the Bernoulli random variable will be used as a building block to form other probability distributions, such as the binomial distribution of Section 4.4. The properties of the Bernoulli distribution are summarized below.

> The Bernoulli Distribution
>
> $$p(x) = p^x(1-p)^{1-x}, \qquad x = 0,\ 1 \qquad \textit{for } 0 \le p \le 1$$
>
> $$E(X) = p \text{ and } V(X) = p(1-p)$$

4.4 The Binomial Distribution

4.4.1 Probability Function

Suppose we conduct n independent Bernoulli trials, each with a probability p of success. Let the random variable Y be the number of successes in the n trials. The distribution of Y is called the binomial distribution. As an illustration, instead of inspecting a single item, as we do with a Bernoulli random variable, suppose that we now independently inspect n items and record values for $X_1, X_2,\ldots,X_n$, where $X_i = 1$ if the ith inspected item is defective and $X_i = 0$, otherwise. The sum of the X_i's,

$$Y = \sum_{i=1}^{n} X_i$$

denotes the number of defectives among the n sampled items.

We can easily find the probability distribution for Y under the assumption that $P(X_i = 1) = p$, where p remains constant over all trials. For the sake of simplicity, let us look at the specific case of $n = 3$. The random variable Y can then take on four possible values: 0, 1, 2, and 3. For Y to be 0, all three X_i values must be 0. Thus,

$$
\begin{aligned}
P(Y = 0) &= P(X_1 = 0,\ \ X_2 = 0,\ \ X_3 = 0) \\
&= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) \\
&= (1-p)^2
\end{aligned}
$$

Now if $Y = 1$, then exactly one value of X_i is 1 and the other two are 0. The one defective could occur on any of the three trials; thus,

$$
\begin{aligned}
P(Y=1) = P[&(X_1=1,\ \ X_2=0,\ \ X_3=0) \cup (X_1=0,\ \ X_2=1,\ \ X_3=0) \\
&\cup (X_1=0,\ \ X_2=0,\ \ X_3=1)] \\
&= P(X_1=1,\ \ X_2=0,\ \ X_3=0) + P(X_1=0,\ \ X_2=1,\ \ X_3=0) \\
&\qquad + P(X_1=0,\ \ X_2=0,\ \ X_3=0) \\
&(\textit{because the three possibilities are mutually exclusive}) \\
&= P(X_1=1)P(X_2=0)P(X_3=0) + P(X_1=0)P(X_2=1)P(X_3=0) \\
&\qquad + P(X_1=0)P(X_2=0)P(X_3=1) \\
&= p(1-p)^2 + p(1-p)^2 + p(1-p)^2 \\
&= 3p(1-p)^2
\end{aligned}
$$

Notice that the probability of each specific outcome is the same, $p(1 - p)^2$.

For $Y = 2$, two values of X_i must be 1 and one must be 0, which can occur in three mutually exclusive ways. Hence,

$$
\begin{aligned}
P(Y=2) = P[&(X_1=1,\ \ X_2=1,\ \ X_3=0) \cup (X_1=1,\ \ X_2=0\ \ X_3=1) \\
&\cup (X_1=0,\ \ X_2=1,\ \ X_3=1)] \\
&= P(X_1=1)P(X_2=0)P(X_3=0) + P(X_1=0)P(X_2=1)P(X_3=0) \\
&\qquad + P(X_1=0)P(X_2=0)P(X_3=1) \\
&= p^2(1-p) + p^2(1-p) + p^2(1-p) \\
&= 3p^2(1-p)
\end{aligned}
$$

The event $Y = 3$ can occur only if all values of X_i are 1, so

$$
\begin{aligned}
P(Y=3) &= P(X_1=1,\ \ X_2=1,\ \ X_3=1) \\
&= P(X_1=1)P(X_2=1)P(X_3=1) \\
&= p^3
\end{aligned}
$$

Notice that the coefficient in each of the expressions for $P(Y = y)$ is the number of ways of selecting y positions, in sequence, in which to place 1's. Because there are three positions in the sequence, this number amounts to $\binom{3}{y}$. Thus we can write

$$
P(Y=y) = \binom{3}{y} p^y (1-p)^{3-y}, \qquad y = 0,\ 1,\ 2,\ 3, \qquad \textit{when}\ n = 3
$$

For general values of n, the probability that Y will take on a specific value—say, y—is given by the term $p^y(1-p)^{n-y}$ multiplied by the number of possible outcomes that result in exactly y defectives being observed. This number, which represents the number of possible ways of selecting y positions for defectives in the n possible positions of the sequence, is given by

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

where $n! = n(n-1)\ldots 1$ and $0! = 1$. Thus, in general, the probability mass function for the binomial distribution is

$$P(Y = y) = p(y) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0,\ 1,\ 2,\ ..,\ n$$

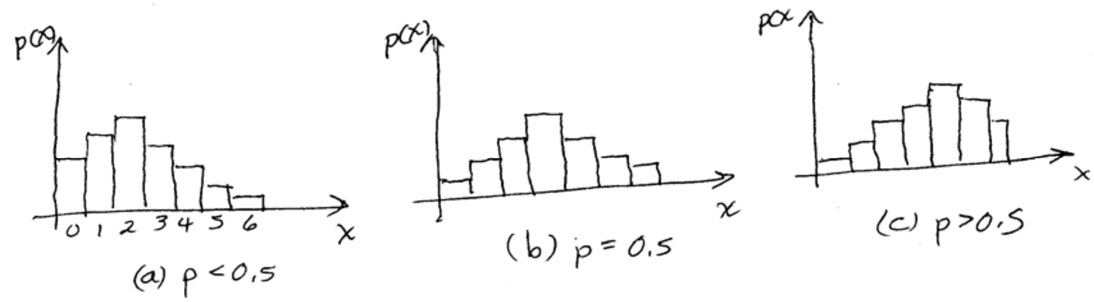
Once n and p are specified we can completely determine the probability function for the binomial distribution; hence, the parameters of the binomial distribution are n and p.

The shape of the binomial distribution is affected by both parameters n and p. If $p = 0.5$, the distribution is symmetric. If $p < 0.5$, the distribution is skewed right, becoming less skewed as n increases. Similarly, if $p > 0.5$, the distribution is skewed left and becomes less skewed as n increases (see Figure 4.6). You can explore the shape of the binomial distribution using the graphing binomial applet. When $n = 1$,

$$p(y) = \binom{1}{y} p^y (1-p)^{1-y} = p^y (1-p)^{1-y}, \qquad y = 0,\ 1$$

the probability function of the Bernoulli distribution. Thus, the Bernoulli distribution is a special case of the binomial distribution with $n = 1$.

Figure 4.6. Binomial probabilities for $p < 0.5$, $p = 0.5$, $p > 0.5$



Notice that the binomial probability function satisfies the two conditions of a probability function. First, probabilities are nonnegative. Second, the sum of the probabilities is one, which can be verified using the binomial theorem:

$$\begin{aligned}\sum_x p(x) &= \sum_{x=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} \\ &= (p + (1-p))^n \\ &= 1\end{aligned}$$

Although we have used $1 - p$ to denote the probability of success, $q = 1 - p$ is a common notation that we will use here and in later sections.

> To summarize, a random variable Y possesses a binomial distribution if the following conditions are satisfied:
> 1. The experiment consists of a fixed number n of identical trials.
> 2. Each trial can result in one of only two possible outcomes, called success or failure; that is, each trial is a Bernoulli trial.
> 3. The probability of success p is constant from trial to trial.
> 4. The trials are independent.
> 5. Y is defined to be the number of successes among the n trials.

Many experimental situations involve random variables that can be adequately modeled by the binomial distribution. In addition to the number of defectives in a sample of n items, examples include the number of employees who favor a certain retirement policy out of n employees interviewed, the number of pistons in an eight-cylinder engine that are misfiring, and the number of electronic systems sold this week out of the n that were manufactured.

Example 4.9:

Suppose that 10% of a large lot of apples are damaged. If four apples are randomly sampled from the lot, find the probability that exactly one apple is damaged. Find the probability that at least one apple in the sample of four is defective.

Solution:

We assume that the four trials are independent and that the probability of observing a damaged apple is the same (0.1) for each trial. This would be approximately true if the lot indeed were *large*. (If the lot contained only a few apples, removing one apple would substantially change the probability of observing a damaged apple on the second draw.) Thus, the binomial distribution provides a reasonable model for this experiment, and we have (with Y denoting the number of defectives)

$$p(1) = \binom{4}{1}(0.1)^1(0.9)^3 = 0.2916$$

To find $P(Y \geq 1)$, we observe that

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) = 1 - p(0) \\ &= 1 - \binom{4}{0}(0.1)^0(0.9)^4 \\ &= 1 - (0.9)^4 \\ &= 0.3439 \end{aligned}$$

Discrete distributions, like the binomial, can arise in situations where the underlying problem involves a continuous (that is, nondiscrete) random variable. The following example provides an illustration.

Example 4.10:

In a study of life lengths for a certain battery for laptop computers, researchers found that the probability that a battery life X will exceed 5 hours is 0.12. If three such batteries are in use in independent laptops, find the probability that only one of the batteries will last 5 hours or more.

Solution:

Letting Y denote the number of batteries lasting 5 hours or more, we can reasonably assume Y to have a binomial distribution, with $p = 0.12$. Hence,

$$P(Y = 1) = p(1) = \binom{3}{1}(0.12)^1(0.88)^2 = 0.279$$

4.4.2 Mean and Variance

There are numerous ways of find $E(Y)$ and $V(Y)$ for a binomially distributed random variable Y. We might use the basic definition and compute

$$\begin{aligned} E(Y) &= \sum_y yp(y) \\ &= \sum_{y=0}^{n} y\binom{n}{y}p^y(1-p)^{n-y} \end{aligned}$$

but direct evaluation of this expression is a bit tricky. Another approach is to make use of the results on linear functions of random variables, which will be presented in Chapter 6. We shall see in Chapter 6 that, because the binomial Y arose as a sum of independent Bernoulli random variables $X_1, X_2, \ldots, X_n$,

$$\begin{aligned} E(Y) &= E\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=0}^{n} E(X_i) \\ &= \sum_{i=1}^{n} p \\ &= np \end{aligned}$$

and

$$V(Y) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Example 4.11:

Referring to example 4.9, suppose that a customer is the one who randomly selected and then purchased the four apples. If an apple is damaged, the customer complains. To keep the customers satisfied, the store has a policy of replacing any damaged item (here the apple) and giving the customer a coupon for future purchases. The cost of this program has, through time, been found to be $C = 3Y^2$, where Y denotes the number of defective apples in the purchase of four. Find the expected cost of the program when a customer randomly selects 4 apples from the lot.

Solution:

We know that

$$E(C) = E\left(3Y^2\right) = 3E(Y^2)$$

and it now remains for us to find $E(Y^2)$. From Theorem 4.3,

$$V(Y) = E(Y-\mu)^2 = E\left(Y^2\right) - \mu^2$$

Since $V(Y) = np(1-p)$ and $\mu = E(Y) = np$, we see that

$$\begin{aligned} E\left(Y^2\right) &= V(Y) + \mu^2 \\ &= np(1-p) + (np)^2 \end{aligned}$$

For example 4.9, $p = 0.1$ and $n = 4$; hence,

$$\begin{aligned} E(C) = 3E\left(Y^2\right) &= 3\left[np(q-p) + (np)^2\right] \\ &= 3\left[4(0.1)(0.9) + (4)^2(0.1)^2\right] \\ &= 1.56 \end{aligned}$$

If the costs were originally expressed in dollars, we could expect to pay an average of \$1.56 when a customer purchases 4 apples.

4.4.3. History and Applications

The binomial expansion can be written as

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

If $a = p$, where $0 < p < 1$, and $b = 1 - p$, we see that the terms on the left are the probabilities of the binomial distribution. Long ago it was found that the binomial coefficients, $\binom{n}{x}$, could be generated from Pascal's triangle (see Figures 4.7 and 4.8).

Figure 4.7. Blaise Pascal (1623—1662)

Source: http://oregonstate.edu/instruct/phl302/philosophers/pascal.html

Figure 4.8. Pascal's triangle

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

To construct the triangle, the first two rows are created, consisting of 1s. Subsequent rows have the outside entries as ones; each of the interior numbers is the sum of the numbers immediately to the left and to the right on the row above.

According to David (1955), the Chinese writer Chu Shih-chieh published the arithmetical triangle of binomial coefficients in 1303, referring to it as an ancient method.

The triangle seems to have been discovered and rediscovered several times. Michael Stifel published the binomial coefficients in 1544 (Eves 1969). Pascal's name seems to have become firmly attached to the arithmetical triangle, becoming known as Pascal's triangle, about 1665, although a triangle-type array was also given by Bernoulli in the 1713 *Ars Conjectandi*.

Jacob Bernoulli (1654-1705) is generally credited for establishing the binomial distribution for use in probability (see Figure 4.9) (Folks 1981, Stigler 1986). Although his father planned for him to become a minister, Bernoulli became interested in and began to pursue mathematics. By 1684, Bernoulli and his younger brother John had developed differential calculus from hints and solutions provided by Leibniz. However, the brothers became rivals, corresponding in later years only be print. Jacob Bernoulli would pose a problem in a journal. His brother John would provide an answer in the same issue, and Jacob would respond, again in print, that John had made an error.

Figure 4.9. Jacob Bernoulli (1654—1705)

Source: http://www.stetson.edu/~efriedma/periodictable/html/Br.html

When Bernoulli died of a "slow fever" on August 16, 1795, he left behind numerous unpublished, and some uncompleted, works. The most important of these was on probability. He had worked over a period of about twenty years prior to his death on the determination of chance, and it was this work that his nephew published in 1713, *Ars Conjectandi* (The Art of Conjecturing). In this work, he used the binomial expansion to address probability problems, presented his theory of permutations and combinations, developed the Bernoulli numbers, and provided the weak law of large numbers for Bernoulli trials.

As was the case with many of the early works in probability, the early developments of the binomial distribution resulted from efforts to address questions relating to games of chance. Subsequently, problems in astronomy, the social sciences, insurance, meteorology, and medicine are but a few of those that have been addressed using this distribution. Polls are frequently reported in the newspaper and on radio and television. The binomial distribution is used to determine how many people to survey and how to present the results. Whenever an event has two possible outcomes and n such events are to be observed, the binomial distribution is generally the first model considered. This has led it to be widely used in quality control in manufacturing processes.

Determining the probabilities of the binomial distribution quickly becomes too complex to be done quickly by hand. Many calculators and software programs have built-in functions for this purpose. Table 2 in the Appendix gives cumulative binomial probabilities for selected values of *n* and *p*. The entries in the table are values of

$$\sum_{y=0}^{a} p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y}, \qquad a = 0,1,\ldots,n-1$$

The following example illustrates the use of the table.

Example 4.12:

An industrial firm supplies ten manufacturing plants with a certain chemical. The probability that any one firm will call in an order on a given day is 0.2, and this probability is the same for all ten plants. Find the probability that, on the given day, the number of plants calling in orders is as follows.

1. at most 3
2. at least 3
3. exactly 3

Solution:

Let *Y* denote the number of plants that call in orders on the day in question. If the plants order independently, then *Y* can be modeled to have a binomial distribution with p = 0.2.

1. We then have

$$\begin{aligned} P(Y \le 3) &= \sum_{y=0}^{3} p(y) \\ &= \sum_{y=0}^{3} \binom{10}{y} (0.2)^y (0.8)^{10-y} \\ &= 0.879 \end{aligned}$$

This may be determined using Table 2(*b*) in the Appendix. Note we use 2(*b*) because it corresponds to *n* = 10. Then the probability corresponds to the entry in column *p* = 0.2 and row *k* = 3. Notice the binomial calculator applet provides the same result.

2. Notice that

$$\begin{aligned} P(Y \ge 3) &= 1 - P(Y \le 2) \\ &= 1 - \sum_{y=0}^{2} \binom{10}{y} (0.2)^y (0.8)^{10-y} \\ &= 1 - 0.678 = 0.322 \end{aligned}$$

Here we took advantage of the fact that positive probability only occurs at integer values so that, for example, $P(Y = 2.5) = 0$.

3. Observe that

$$P(Y = 3) = P(Y \le 3) - P(Y \le 2)$$
$$= 0.879 - 0.678 = 0.201$$

from the results just established.

The examples used to this point have specified n and p in order to calculate probabilities or expected values. Sometimes, however, it is necessary to choose n so as to achieve a specified probability. Example 4.13 illustrates the point.

Example 4.13:

Every hospital has backup generators for critical systems should the electricity go out. Independent but identical backup generators are installed so that the probability that at least one system will operate correctly when called upon is no less than 0.99. Let n denote the number of backup generators in a hospital. How large must n be to achieve the specified probability of at least one generator operating, if

1. $p = 0.95$?
2. $p = 0.8$?

Solution:

Let Y denote the number of correctly operating generators. If the generators are identical and independent, Y has a binomial distribution. Thus,

$$P(Y \ge 1) = 1 - P(Y = 0)$$
$$= 1 - \binom{n}{0} p^0 (1-p)^n$$
$$= 1 - (1-p)^n$$

The conditions specify that n must be such that $P(Y \ge 1) = 0.99$ or more.

1. When p =0.95,

$$P(Y \ge 1) = 1 - (1 - 0.95)^n \ge 0.99$$

results in

$$1 - (0.05)^n \ge 0.99$$

or

$$(0.05)^n \le 1 - 0.99 = 0.01$$

so $n = 2$; that is, installing two backup generators will satisfy the specifications.

2. When $p = 0.80$,

$$P(Y \ge 1) = 1 - (1 - 0.8)^n \ge 0.99$$

results in

$$(0.2)^n \le 0.01$$

Now $(0.2)^2 = 0.04$, and $(0.2)^3 = 0.008$, so we must go to $n = 3$ systems to ensure that

$$P(Y \ge 1) = 1 - (0.2)^3 = 0.992 > 0.99$$

Note: We cannot achieve the 0.99 probability exactly, because Y can assume only integer values.

Example 4.14:

Virtually any process can be improved by the use of statistics, including the law. A much-publicized case that involved a debate about probability was the Collins case, which began in 1964. An incident of purse snatching in the Los Angeles area led to the arrest of Michael and Janet Collins. At their trial, an "expert" presented the following probabilities on characteristics possessed by the couple seen running from the crime. The chance that a couple had all of these characteristics together is 1 in 12 million. Since the Collinses had all of the specified characteristics, they must be guilty. What, if anything, is wrong with this line of reasoning?

Man with beard	$\frac{1}{10}$
Blond woman	$\frac{1}{4}$
Yellow car	$\frac{1}{10}$
Woman with ponytail	$\frac{1}{10}$
Man with mustache	$\frac{1}{3}$
Interracial couple	$\frac{1}{1000}$

Solution:

First, no background data are offered to support the probabilities used. Second, the six events are not independent of one another and, therefore, the probabilities cannot be multiplied. Third, and most interesting, the wrong question is being addressed. The question of interest is not "What is the probability of finding a couple with these characteristics?" Since one such couple has been found (the Collinses), the proper question is: "What is the probability that *another* such couple exists, given that we found one?" Here is where the binomial distribution comes into play. In the binomial model, let

n = Number of couples who could have committed the crime
p = Probability that any one couple possesses the six listed characteristics
x = Number of couples who possess the six characteristics
From the binomial distribution, we know that

$$P(X=0)=(1-p)^n$$
$$P(X=1)=np(-p)^{n-1}$$
$$P(X\geq 1)=1-(1-p)^n$$

Then, the answer to the conditional question posed above is

$$\begin{aligned} P(X>1\mid X\geq 1) &= \frac{P[(X>1)\cap(X\geq 1)]}{P(X\geq 1)} \\ &= \frac{P(X>1)}{P(X\geq 1)} \\ &= \frac{1-(1-p)^n - np(1-p)^{n-1}}{1-(1-p)^n} \end{aligned}$$

Substituting p = 1/12 million and n = 12 million, which are plausible but not well-justified guesses, we get

$$P(X>1\mid X\geq 1)=0.42$$

so the probability of seeing another such couple, given that we have already seen one, is much larger than the probability of seeing such a couple in the first place. This holds true even if the numbers are dramatically changed. For instance, if n is reduced to 1 million, the conditional probability becomes 0.05, which is still much larger than 1/12 million.

The important lessons illustrated here are that the correct probability question is sometimes difficult to determine and that conditional probabilities are *very* sensitive to conditions.

We shall soon move on to a discussion of other discrete random variables; but the binomial distribution, summarized below, will be used frequently throughout the remainder of the text.

The Binomial Distribution

$$p(y)=\binom{n}{y}p^y(1-p)^{n-y}, \qquad y=0,\ 1,\ 2,\ ..,\ n \qquad for\ 0\leq p\leq 1$$

$$E(Y)=np \qquad\qquad V(Y)=np(1-p)$$

Exercises

4.27. Let X denote a random variable that has a binomial distribution with p = 0.3 and n = 5. Find the following values.

a. $P(X=3)$

b. $P(X\leq 3)$

c. $P(X \geq 3)$
d. $E(X)$
e. $V(X)$

4.28. Let X denote a random variable that has a binomial distribution with $p = 0.6$ and $n = 25$. Use your calculator, Table 2 in the Appendix, or the binomial calculator applet to evaluate the following probabilities.
a. $P(X \leq 10)$
b. $P(X \geq 15)$
c. $P(X = 10)$

4.29. A machine that fills milk cartons underfills a certain proportion p. If 50 boxes are randomly selected from the output of this machine, find the probability that no more than 2 cartons are underfilled when
a. $p = 0.05$
b. $p = 0.1$

4.30. When testing insecticides, the amount of the chemical when, given all at once, will result in the death of 50% of the population is called the LD50, where LD stands for lethal dose. If 40 insects are placed in separate Petri dishes and treated with an insecticide dosage of LD50, find the probabilities of the following events.
a. Exactly 20 survive
b. At most 15 survive
c. At least 20 survive

4.31. Refer to Exercise 4.30.
a. Find the number expected to survive, out of 40.
b. Find the variance of the number of survivors, out of 40.

4.32. Among persons donating blood to a clinic, 85% have Rh^+ blood (that is, the Rhesus factor is present in their blood.) Six people donate blood at the clinic on a particular day.
a. Find the probability that at least one of the five does not have the Rh factor.
b. Find the probability that at most four of the six have Rh^+ blood.

4.33. The clinic in Exercise 4.32 needs six Rh^+ donors on a certain day. How many people must donate blood to have the probability of obtaining blood from at least six Rh^+ donors over 0.95?

4.34. During the 2002 Survey of Business Owners (SBO), it was found that the numbers of female-owned, male-owned, and jointly male- and female-owned business were 6.5, 13.2, and 2.7 million, respectively. Among four randomly selected businesses, find the probabilities of the following events.
a. All four had a female, but no male, owner.
b. One of the four was either owned or co-owned by a male.
c. None of the four were jointly owned by female and a male.

4.35. Goranson and Hall (1980) explain that the probability of detecting a crack in an airplane wing is the product of p_1, the probability of inspecting a plane with a wing crack; p_2, the probability of inspecting the detail in which the crack is located; and p_3, the probability of detecting the damage.
a. What assumptions justify the multiplication of these probabilities?
b. Suppose that $p_1 = 0.9$, $p_2 = 0.8$, and $p_3=0.5$ for a certain fleet of planes. If three planes are inspected from this fleet, find the probability that a wing crack will be detected in at least one of them.

4.36. Each day a large animal clinic schedules 10 horses to be tested for a common respiratory disease. The cost of each test is $80. The probability of a horse having the disease is 0.1. If the horse has the disease, treatment costs $500.
a. What is the probability that at least one horse will be diagnosed with the disease on a randomly selected day?
b. What is the expected daily revenue that the clinic earns from testing horses for the disease and treating those who are sick?

4.37. The efficacy of the mumps vaccine is about 80%; that is, 80% of those receiving the mumps vaccine will not contract the disease when exposed. Assume each person's response to the mumps is independent of another person's response. Find the probability that at least one exposed person will get the mumps if n are exposed where
a. $n = 2$
b. $n = 4$

4.38. Refer to Exercise 4.37.
a. How many vaccinated people must be exposed to the mumps before the probability that at least one person will contract the disease is at least 0.95?
b. In 2006, an outbreak of mumps in Iowa resulted in 605 suspect, probable, and confirmed cases. Given broad exposure, do you find this number to be excessively large? Justify your answer.

4.39. A complex electronic system is built with a certain number of backup components in its subsystems. One subsystem has four identical components, each with a probability of 0.15 of failing in less than 1000 hours. The subsystem will operate if any two or more of the four components are operating. Assuming that the components operate independently, find the probabilities of the following events.
a. Exactly two of the four components last longer than 1000 hours.
b. The subsystem operates for longer than 1000 hours.

4.40. In a study, dogs were trained to detect the presence of bladder cancer by smelling urine (*USA Today*, September 24, 2004). During training, each dog was presented with urine specimens from healthy people, those from people with bladder cancer, and those from people sick with unrelated diseases. The dog was trained to lie down by any urine specimen from a person with bladder cancer. Once training was completed, each dog was presented with seven urine specimens, only one of which came from a person with

bladder cancer. The specimen that the dog laid down beside was recorded. Each dog repeated the test nine times. Six dogs were tested.
a. One dog had only one success in 9. What is the probability of the dog having at least this much success if it cannot detect the presence of bladder cancer by smelling a person's urine?
b. Two dogs correctly identified the bladder cancer specimen on 5 of the 9 trials. If neither were able to detect the presence of bladder cancer by smelling a person's urine, what is the probability that both dogs correctly detected the bladder specimen on at least 5 of the 9 trials?

4.41. A firm sells four items randomly selected from a large lot that is known to contain 12% defectives. Let Y denote the number of defectives among the four sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by

$$C = 2Y^2 + Y + 3$$

Find the expected repair cost.

4.42. From a large lot of memory chips for use in personal computes, n are to be sampled by a potential buyer, and the number of defectives X is to be observed. If at least one defective is observed in the sample of n, the entire lot is to be rejected by the potential buyer. Find n so that the probability of detecting at least one defective is approximately 0.95 if the following percentages are correct.
a. 10% of the lot is defective.
b. 5% of the lot is defective.

4.43. Fifteen free-standing ranges with smoothtops are available for sale in a wholesale appliance dealer's warehouse. The ranges sell for \$550 each, but a double-your-money-back guarantee is in effect for any defective range the purchaser might purchase. Find the expected net gain for the seller if the probability of any one range being defective is 0.06. (Assume that the quality of any one range is independent of the quality of the others.)

4.5 The Geometric Distribution

4.5.1 Probability Function

Suppose that a series of test firings of a rocket engine can be represented by a sequence of independent Bernoulli random variables, with $X_i = 1$ if the ith trial results in a successful firing and with $X_i = 0$, otherwise. Assume that the probability of a successful firing is constant for the trials, and let this probability be denoted by p. For this problem, we might be interested in the number of failures prior to the trial on which the first successful firing occurs. If Y denotes the number of failures prior to the first success, then

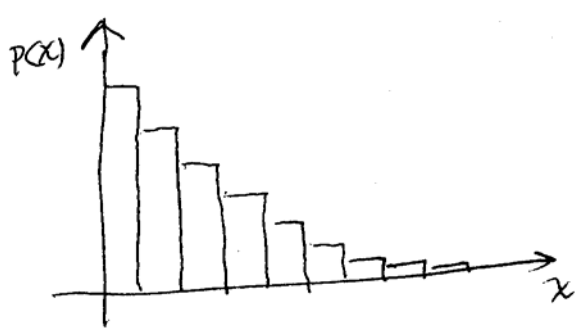
$$
\begin{aligned}
P(Y=y) = p(y) &= P(X_1 = 0, X_2 = 0, \ldots, X_y = 0, X_{y+1} = 1) \\
&= P(X_1 = 0)P(X_2 = 0)\cdots P(X_y = 0)P(X_{y+1} = 1) \\
&= (1-p)(1-p)\cdots(1-p)p \\
&= (1-p)^y p, \\
&= q^y p \qquad\qquad y = 0,\ 1,\ 2,\ \ldots
\end{aligned}
$$

because of the independence of the trials. This formula is referred to as the geometric probability distribution. Notice that this random variable can take on a countably infinite number of possible values. In addition,

$$
\begin{aligned}
P(Y=y) &= q^y p \\
&= q\left[q^{y-1}p\right] \\
&= qP(Y = y-1) \\
&< P(Y = y-1), \qquad y = 1,\ 2,\ \ldots
\end{aligned}
$$

That is, each succeeding probability is less than the previous one (see Figure 4.10).

Figure 4.10. Geometric distribution probability function



In addition to the rocket-firing example just given, other situations may result in a random variable whose probability can be modeled by a geometric distribution: the number of customers contacted before the first sale is made; the number of years a dam is in service before it overflows; and the number of automobiles going through a radar check before the first speeder is detected. The following example illustrates the use of the geometric distribution.

Example 4.15:

A recruiting firm finds that 20% of the applicants for a particular sales position are fluent in both English and Spanish. Applicants are selected at random from the pool and interviewed sequentially. Find the probability that five applicants are interviewed before finding the first applicant who is fluent in both English and Spanish.

Solution:

Each applicant either is or is not fluent in English and Spanish, so the interview of an applicant corresponds to a Bernoulli trial. The probability of finding a suitable applicant will remain relatively constant from trial to trial if the pool of applicants is reasonably large. Because applicants will be interviewed until the first one fluent in English and Spanish is found, the geometric distribution is appropriate. Let Y = the number of *unqualified* applicants prior to the first qualified one. If five unqualified applicants are interviewed before finding the first applicant who is fluent in English and Spanish, we want to find the probability that $Y = 5$. Thus,

$$\begin{aligned} P(Y=5) &= p(5) = (0.8)^5(0.2) \\ &= 0.066 \end{aligned}$$

The name of the geometric distribution comes from the geometric series its probabilities represent. Properties of the geometric series are useful when finding the probabilities of the geometric distribution. For example, the sum of a geometric series is

$$\sum_{x=0}^{\infty} t^x = \frac{1}{1-t}$$

for |t| <1. Using this fact, we can show the geometric probabilities sum to one:

$$\begin{aligned} \sum_y p(y) &= \sum_{y=0}^{\infty}(1-p)^y p \\ &= p\sum_{y=0}^{\infty}(1-p)^y \\ &= p\frac{1}{1-(1-p)} \\ &= 1 \end{aligned}$$

Similarly, using the partial sum of a geometric series, we can find the functional form of the geometric distribution function. For any integer $y \geq 0$,

$$\begin{aligned} F(y) = P(Y \leq y) &= \sum_{t=0}^{y} q^t p \\ &= p\sum_{t=0}^{y} q^t \\ &= p\frac{1-q^{y+1}}{1-q} \\ &= p\frac{1-q^{y+1}}{p} \\ &= 1-q^{y+1} \end{aligned}$$

Using the distribution function, we have, for any integer $y \geq 0$,

$$P(Y > y) = 1 - F(y) = 1 - (1 - q^{y+1}) = q^{y+1}$$

4.5.2. Mean and Variance

From the basic definition,

$$\begin{aligned} E(Y) &= \sum_{y} yp(y) = \sum_{y=0}^{\infty} ypq^{y} \\ &= p\sum_{y=0}^{\infty} yq^{y} \\ &= p\left[0 + q + 2q + 3q^{2} + \cdots\right] \\ &= p\,q\left[1 + 2q + 3q^{2} + \cdots\right] \end{aligned}$$

The infinite series can be split up into a triangular array of series as follows:

$$\begin{aligned} E(Y) = pq[1 + q + q^{2} + \cdots \\ + q + q^{2} + \cdots \\ + q^{2} + \cdots \\ + \cdots] \end{aligned}$$

Each line on the right side is an infinite, decreasing geometric progression with common ratio q. Recall that $a + ax + ax^{2} + \cdots = a/(1 - x)$ if $|x| < 1$. Thus, the first line inside the bracket sums to $1/(1 - q) = 1/p$; the second, to q/p; the third, to q^2/p; and so on. On accumulating these totals, we then have

$$\begin{aligned} E(Y) &= pq\left[\frac{1}{p} + \frac{q}{p} + \frac{q^{2}}{p} + \cdots\right] \\ &= q\left[1 + q + q^{2} + \cdots\right] \\ &= \frac{q}{1 - q} \\ &= \frac{q}{p} \end{aligned}$$

This answer for $E(Y)$ should seem intuitively realistic. For example, if 10% of a certain lot of items are defective, and if an inspector looks at randomly selected items one at a time, she should expect to find nine good items before finding the first defective one.

The variance of the geometric distribution will be derived in Section 4.4 and in Chapter 6. The result, however, is

$$V(Y) = \frac{q}{p^2}$$

Example 4.16:

Referring to example 4.15, let Y denote the number of unqualified applicants interviewed prior to the first qualified one. Suppose that the first applicant fluent in both English and Spanish is offered the position, and the applicant accepts. If each interview costs $125, find the expected value and the variance of the total cost of interviewing until the job is filled. Within what interval should this cost be expected to fall?

Solution:

Because $(Y + 1)$ is the number of the trial on which the interviewing process ends, the total cost of interviewing is $C = 125(Y + 1) = 125Y + 125$. Now,

$$\begin{aligned} E(C) &= 125E(Y) + 125 \\ &= 125\left(\frac{q}{p}\right) + 125 \\ &= 125\left(\frac{0.8}{0.2}\right) + 125 \\ &= 625 \end{aligned}$$

and

$$\begin{aligned} V(C) &= (125)^2 V(Y) \\ &= 125\left(\frac{q}{p^2}\right) \\ &= 125\left(\frac{0.8}{(0.2)^2}\right) \\ &= 2500 \end{aligned}$$

The geometric is the only discrete distribution that has the *memoryless property*. By this we mean that, if we know the number of failures exceed j, the probability that there will be more than $j + k$ failures prior to the first success is equal to the probability that the number of failures exceeds k; that is, for integers j and k greater than 0,

$$P(Y > j + k \mid Y > j) = P(Y > k)$$

To verify this, we will use the properties of conditional probabilities.

$$
\begin{aligned}
P(Y > j+k \mid Y > j) &= \frac{P\big((Y > j+k) \cap (Y > j)\big)}{P(Y > j)} \\
&= \frac{P(Y > j+k)}{P(Y > j)} \\
&= \frac{q^{j+k}}{q^{j}} \\
&= q^{k} \\
&= P(Y > k)
\end{aligned}
$$

Example 4.17:

Referring once again to Example 4.15, suppose that 10 applicants have been interviewed and no person fluent in both English and Spanish have been identified. What is the probability that 15 unqualified applicants will be interviewed before finding the first applicant who is fluent in English and Spanish?

Solution:

By the memoryless property, the probability that fifteen unqualified applicants will be interviewed before finding an applicant who is fluent in English and Spanish, given the first ten are not qualified is equal to the probability of finding the first qualified candidate after interviewing 5 unqualified applicants. Again, let Y denote the number of unqualified applicants interviewed prior to the first candidate who is fluent in English and Spanish. Thus,

$$
\begin{aligned}
P(Y = 15 \mid Y \geq 10) &= \frac{P\big((Y = 15) \cap (Y \geq 10)\big)}{P(Y \geq 10)} \\
&= \frac{P(Y = 15)}{P(Y > 9)} \\
&= \frac{pq^{15}}{q^{10}} \\
&= pq^{5} \\
&= P(Y = 5)
\end{aligned}
$$

The Geometric Distribution

$$p(y) = p(1-p)^y, \qquad y = 0,\ 1,\ 2,\ .. \qquad for\ 0 \le p \le 1$$

$$E(Y) = \frac{q}{p} \qquad\qquad V(Y) = \frac{q}{p^2}$$

4.6 The Negative Binomial Distribution

4.6.1. Probability Function

In section 4.5, we saw that the geometric distribution models the probabilistic behavior of the number of failures prior to the *first success* in a sequence of independent Bernoulli trials. But what if we were interested in the number of failures prior to the second success, or the third success, or (in general) the rth success. The distribution governing probabilistic behavior in these cases is called the *negative binomial distribution*.

Let Y denote the number of failures prior to the rth success in a sequence of independent Bernoulli trials, with p denoting the common probability of success. We can derive the distribution of Y from known facts. Now,

P($Y = y$) = P(1^{st} ($y + r – 1$) trials contain ($r – 1$) successes and the ($y + r$)th trial is a success)
= P[1^{st} y trials contain ($r – 1$) successes] x P[yth trial is a success]

Because the trials are independent, the joint probability can be written as a product of probabilities. The first probability statement is identical to the one that results in a binomial model; and hence,

$$\begin{aligned} P(Y = y) &= p(y) \\ &= \binom{y+r-1}{r-1} p^{r-1}(1-p)^y \times p \\ &= \binom{y+r-1}{r-1} p^r q^y, \qquad y = 0,\ 1,\ ... \end{aligned}$$

Notice that, if $r = 1$, we have the geometric distribution. Thus, the geometric distribution is a special case of the negative binomial distribution. The negative binomial is a quite flexible model. Its shape ranges from one being highly skewed to the right when r is small to one that is relatively symmetric as r becomes large and p small. Some of these are displayed in Figure 4.11.

Figure 4.11. Form of the negative binomial distribution

Example 4.18:

As in Example 4.15, 20% of the applicants for a certain sales position are fluent in English and Spanish. Suppose that four jobs requiring fluency in English and Spanish are open. Find the probability that two unqualified applicants are interviewed before finding the fourth qualified applicant, if the applicants are interviewed sequentially and at random.

Solution:

Again we assume independent trials, with 0.2 being the probability of finding a qualified candidate on any one trial. Let Y denote the number of unqualified applicants interviewed prior to interviewing the 4$^{\text{th}}$ applicant who is fluent in English and Spanish. Y can reasonably be assumed to have a negative binomial distribution, so

$$\begin{aligned} P(Y=2) = p(2) &= \binom{5}{3}(0.2)^4(0.8)^2 \\ &= 10(0.2)^4(0.8)^2 \\ &= 0.01 \end{aligned}$$

4.6.2 Mean and Variance

The expected value, or mean, and the variance for the negative binomially distributed random variable Y can easily be found by analogy with the geometric distribution. Recall that Y denotes the number of the failures prior to the rth success. Let W_1 denote the number of failures prior to the first success; let W_2 denote the number of failures between the first success and the second success; let W_3 denote the number of failures between the second success and the third success; and so forth. The results of the trials can then be represented as follows (where F stands for failure and S represents a success):

$$\underbrace{FF\cdots F}_{W_1}\underline{S},\underbrace{FF\cdots}_{W_2}\underline{S},\underbrace{FF\cdots F}_{W_3}\underline{S}$$

Clearly, $Y = \sum_{i=1}^{r} W_i$, where the W_i values are independent and each has a geometric distribution. Thus, by results to be derived in Chapter 6,

$$E(Y) = \sum_{i=1}^{r} E(W_i) = \sum_{i=1}^{r} \left(\frac{q}{p} \right) = \frac{rq}{p}$$

and

$$V(Y) = \sum_{i=1}^{r} V(W_i) = \sum_{i=1}^{r} \left(\frac{q}{p^2} \right) = \frac{rq}{p^2}$$

Example 4.19:

A swimming pool repairman has three check valves in stock. Ten percent of the service calls require a check valve. What is the expected number and variance of the number of service calls she will make before using the last check valve?

Solution:

Let Y denote the number of service calls that do not require a check valve that the repairman will make before using the last check valve. Assuming that each service call is independent from the others and that the probability of needing a check valve is constant for each service call, the negative binomial is a reasonable model with $r = 3$ and $p = 0.1$. The total number of service calls made to use the three check valves is $C = Y + 3$. Now,

$$\begin{aligned} E(C) &= E(Y) + 3 \\ &= \frac{3(0.9)}{0.1} + 3 \\ &= 30 \end{aligned}$$

and

$$\begin{aligned} V(C) &= V(Y) \\ &= \frac{3(0.9)}{0.1^2} \\ &= 270 \end{aligned}$$

4.6.3. History and Applications

William S. Gosset (1876-1937) studied mathematics and chemistry at New College Oxford before joining the Arthur Guinness Son and Company in 1899 (see Figure 4.12). At the brewery, he worked on a variety of mathematical and statistical problems that arose in the brewing process, publishing the results of his efforts under the pen name "Student." He encountered the negative binomial while working with the distributions of yeast cells counted with a haemocytometer (1907). Gosset reasoned that, if the liquid in which the cells were suspended was properly mixed, then a given particle had an equal chance of falling on any unit area of the haemocytometer. Thus, he was working with the binomial distribution, and he focused on estimating the parameters n and p. To his surprise, in two of his four series, the estimated variance exceeded the mean, resulting in negative estimates of n and p. Nevertheless, these "negative" binomials fit his data well. He noted that this may have occurred due to a tendency of the yeast cells "to stick together in groups which was not altogether abolished even by vigorous shaking" (p. 357).

Several other cases appeared in the literature during the early 1900's where estimation of the binomial parameters resulted in negative values of p and n. This phenomenon was explained to some extent by arguing that for small p and large n, the variability of the estimators would cause some negative estimates to be observed. Whitaker (1915) investigated the validity of this claim. In addition to Student's work, she reviewed that of Mortara who dealt with deaths due to chronic alcoholism and that of Bortkewitsch who studied suicides of children in Prussia, suicides of women in German states, accidental deaths in trade societies, and deaths from the kick of a horse in Prussian army corps. Whitaker found it highly unlikely that all negative estimates of p and n could be explained by variability. She, therefore, suggested that a new interpretation was needed with the negative binomial distribution.

Although we have motivated the negative binomial distribution as being the number of failures prior to the rth success in independent Bernoulli trials, several other models have been described that gives rise to this distribution (Boswell and Patil 1970). The Pólya distribution and the Pascal distribution have been other names for the negative binomial distribution. The Pólya distribution was motivated from an urn model and generally refers to the special case where r is a positive integer, although r may be any positive number for the more general negative binomial distribution.

Figure 4.12. William S. Gosset

Source: http://www.gap-system.org/~history/PictDisplay/Gosset.html

Because the negative binomial distribution has been derived in so many different ways, it has also been presented using different parameterizations as well. Some define the negative binomial as the number of *trials* required to get the first success. Because we must have at least one trial, the points of positive probability begin with 1, not 0, as presented earlier. The negative binomial has been used extensively to model the number of organisms within a sampling unit. For these applications, the parameters are often taken to be *r* and the mean μ, instead of *r* and *p*, because the mean μ is of primary interest in such studies. For these reasons, it is important to read carefully how the negative binomial random variable is defined when going from one source to another.

The negative binomial distribution has been applied in many fields including accident statistics, population counts, psychological data, and communications. Some of these applications will be highlighted in the exercises.

The Negative Binomial Distribution

$$p(y) = \binom{y+r-1}{r-1} p^r (1-p)^y, \qquad y = 0,\ 1,\ 2,\ .. \qquad for\ 0 \le p \le 1$$

$$E(Y) = \frac{rq}{p} \qquad\qquad V(Y) = \frac{rq}{p^2}$$

Exercises

4.44. Let Y denote a random variable that has a geometric distribution, with a probability of success on any trial denoted by p. Let $p = 0.1$.

a. Find $P(Y > 2)$.
b. Find $P(Y > 4|Y > 2)$

4.45. Let Y denote a negative binomial random variable, with $p = 0.5$. Find $P(Y \geq 4)$ for the following values of r.
a. $r = 2$
b. $r = 4$

4.46. Suppose that 10% of the engines manufactured on a certain assembly line are defective. If engines are randomly selected one at a time and tested, find the probability that two defective engines will be found before a good engine is found.

4.47. Referring to Exercise 4.46, find the probability that the 5^{th} nondefective engine will be found as follows:
a. After obtaining 2 defectives
b. After obtaining 4 defectives

4.48. Referring to Exercise 4.46, given that the first two engines are defective, find the probability that at least two more defecties are tested before the first nondefective engine is found.

4.49. Referring to Exercise 4.46, find the mean and the variance of the number of defectives tested before the following events occur.
a. The first nondefective engine is found
b. The third nondefective engine is found

4.50. Greenbugs are pests in oats. If their populations get too high, the crop will be destroyed. When recording the number of greenbugs on randomly selected seedling oat plants, the counts have been found to be modeled well by the geometric distribution. Suppose the average number of greenbugs on a seedling oat plant is one. Find the probability that a randomly selected plant has
a. No greenbugs
b. Two greenbugs
c. At least one greenbug

4.51. The employees of a firm that does asbestos cleanup are being tested for indications of asbestos in their lungs. The firm is asked to send four employees who have positive indications of asbestos on to a medical center for further testing. If 40% of the employees have positive indications of asbestos in their lungs, find the probability that six employees who do not have asbestos in their lungs must be tested before finding the four who do have asbestos in their lungs.

4.52. Referring to Exercise 4.51, if each test costs \$40, find the expected value and the variance of the total cost of conducting the tests to locate four positives. Is it highly likely that the cost of completing these tests will exceed \$650?

4.53. People with O^{-} negative blood are called universal donors because they may give blood to anyone without risking incompatibility due to the Rh factor. Nine percent of the persons donating blood at a clinic have O^{-} negative blood, find the probabilities of the following events.
a. The first O^{-} negative donor is found after blood typing 5 people who were not O^{+}
b. The second O^{-} negative donor is the sixth donor of the day.

4.54. A geological study indicates that an exploratory oil well drilled in a certain region should strike oil with probability 0.25. Find the probabilities of the following events.
a. The first strike of oil comes after drilling three dry (non-productive) wells
b. The third strike of oil comes after three dry wells
c. What assumptions must be true for your answers to be correct?

4.55. In the setting of Exercies 4.51, suppose that a company wants to set up three producing wells. Find the expected value and the variance of the number of wells that must be drilled to find three successful ones. (Hint: First find the expected value and variance of the number of dry wells that will be drilled before finding the three successful ones.)

4.56. A large lot of tires contains 5% defectives. Four tires are to be chosen from the lot and placed on a car.
a. Find the probability that two defectives are found before four good ones
b. Find the expected value and the variance of the number of selections that must be made to get four good tires. (Hint: First find the expected value and variance of the number of defective tires that will be selected before finding the four good ones.)

4.57. An interviewer is given a list of potential people she can interview. Suppose that the interviewer needs to interview 5 people and that each person independently agrees to be interviewed with probability 0.6. Let X be the number of people she must ask to be interviewed to obtain her necessary number of interviews.
a. What is the probability that she will be able to obtain the 5 people by asking no more than 7 people?
b. What is the expected value and variance of the number of people she must ask to interview 5 people?

4.58. A car salesman is told that he must make three sales each day. The salesman believes that, if he visits with a customer, the probability that customer will purchase a car is 0.2.
a. What is the probability that the salesman will have to visit with at least five customers to make three sales?
b. What is the expected number of customers the salesman must visit with to make his daily sales goal?

4.59. The number of cotton fleahoppers (a pest) on a cotton plant has been found to be modeled well using a negative binomial distribution with $r = 2$. Suppose the average

number of cotton fleahoppers on plants in a cotton field is two. Find the probability that a randomly selected cotton plant has the following number of fleahoppers.
a. No cotton fleahopper
b. 5 cotton fleahoppers
c. At least one cotton fleahopper

4.60. The number of thunderstorm days in a year has been modeled using a negative binomial model (Sakamoto 1973). A thunderstorm day is defined as a day during which at least one thunderstorm cloud (cumulonimbus) occurs accompanied by lightning and thunder. It may or may not be accompanied by strong gusts of wind, rain, or hail. For one such site, the mean and variance of the number thunderstorm days are 25 days and 40 days^2, respectively. For a randomly selected year, find the probabilities of the following events.
a. No thunderstorm days during a year
b. 20 thunderstorm days during a year
c. At least 40 thunderstorm days during a year

4.61. Refer again to Exercise 4.60. Someone considering moving to the area is concerned about the number of thunderstorm days in a year. He wants assurances that there will be only a 10% chance of the number of thunderstorm days exceeding a specified number of days. Find the number of days that you may properly use in making this assurance.

4.62. This problem is known as the Banach Match Problem. A pipe-smoking mathematician always caries two matchboxes, one in his right-hand pocket and one in his left-hand pocket. Each time he needs a match he is equally likely to take it from either pocket. The mathematician discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches in the other box, $k = 0, 1, 2, \ldots, N$?

4.63. Refer to Exercise 4.62. Suppose that instead of N matches in each box, the left-hand pocket originally has N_1 matches and the one in the right-hand pocket originally has N_2 matches. What is the probability that there are exactly k matches in the other box, $k = 0, 1, 2, \ldots, N$?

4.7. The Poisson Distribution

4.7.1 Probability Function

A number of probability distributions come about through limiting arguments applied to other distributions. One very useful distribution of this type is called the *Poisson distribution*.

Consider the development of a probabilistic model for the number of accidents that occur at a particular highway intersection in a period of one week. We can think of the time interval as being split up into n subintervals such that

P(One accident in a subinterval) = p
P(No accidents in a subinterval) = $1 - p$

Here we are assuming that the same value of p holds for all subintervals, and that the probability of more than one accident occurring in any one subinterval is zero. If the occurrence of accidents can be regarded as independent from subinterval to subinterval, the total number of accidents in the time period (which equals the total number of subintervals that contain one accident) will have a binomial distribution.

Although there is no unique way to choose the subintervals—and we therefore know neither n nor p—it seems reasonable to assume that, as n increases, p should decrease. Thus, we want to look at the limit of the binomial probability distribution as $n \to \infty$ and $p \to 0$. To get something interesting, we take the limit under the restriction that the mean (np in the binomial case) remains constant at a value we will call λ.

Now with $np = \lambda$ or $p = \lambda/n$, we have

$$\lim_{n\to\infty}\binom{n}{y}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \lim_{n\to\infty}\frac{\lambda^{y}}{y!}\left(1-\frac{\lambda}{n}\right)^{n}\frac{n(n-1)\cdots(n-y+1)}{n^{y}}\left(1-\frac{\lambda}{n}\right)^{-y}$$

$$= \lim_{n\to\infty}\frac{\lambda^{y}}{y!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-y}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{y-1}{n}\right)$$

Noting that

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

and that all other terms involving n tend toward unity, we have the limiting distribution

$$p(y) = \frac{\lambda^{y}}{y!}e^{-\lambda}, \qquad y = 0,\ 1,\ 2,\ .. \qquad \textit{for } \lambda > 0$$

Recall that λ denotes the mean number of occurrences in one time period (a week, for the example under consideration); hence, if t non-overlapping time periods were considered, the mean would be λt. Based on this derivation, the Poisson distribution is often referred to as the distribution of rare events.

This distribution, called the *Poisson distribution with parameter* λ, can be used to model counts in areas of volumes, as well as in time. For example, we may use this distribution to model the number of flaws in a square yard of textile, the number of bacteria colonies in a cubic centimeter of water, or the number of times a machine fails in the course of a workday. We illustrate the use of the Poisson distribution in the following example.

Example 4.19:

During business hours, the number of calls passing through a particular cellular relay system averages five per minute. Find the probability that no call wil be received during a given minute.

Solution:

If calls tend to occur independently of one another, and if they occur at a constant rate over time, the Poisson model provides an adequate presentation of the probabilties. Thus,

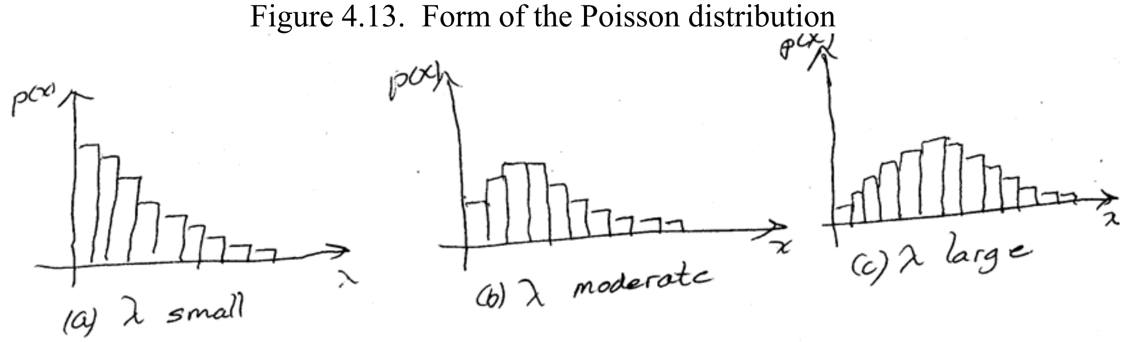
$$p(0) = \frac{5^0}{0!} e^{-5} = e^{-5} = 0.007$$

The shape of the Poisson changes from highly skewed when λ is small to fairly symmetric when λ is large. (See Figure 4.13) Calculators and functions in computer software often can be used to find the probabilities from the probability mass function and the cumulative distribution function. Table 3 of the Appendix gives values for cumulative Poisson probabilities of the form

$$\sum_{y=0}^{a} e^{-\lambda} \frac{\lambda^y}{y!}, \qquad a = 0,\ 1,\ 2, \ldots$$

for selected values of λ.

Figure 4.13. Form of the Poisson distribution



The following example illustrates the use of Table 3 and the Poisson applet.

Example 4.20:

Refer to Example 4.19, and let Y denote the number of calls in the given minute. Find $P(Y \leq 4)$, $P(Y \geq 4)$, and $P(Y = 4)$.

Solution:

From Table 3 or the Poisson applet,

$$P(Y \le 4) = \sum_{y=0}^{4} \frac{(5)^y}{y!} e^{-5} = 0.440$$

Also,

$$\begin{aligned} P(Y \ge 4) &= 1 - P(Y \le 3) \\ &= 1 - \sum_{y=0}^{3} \frac{(5)^y}{y!} e^{-5} \\ &= 1 - 0.265 = 0.735 \end{aligned}$$

and

$$\begin{aligned} P(Y = 4) &= P(Y \le 4) - P(Y \le 3) \\ &= 0.440 - 0.265 \\ &= 0.175 \end{aligned}$$

4.7.2. Mean and Variance

We can intuitively determine what the mean and the variance of a Poisson distribution should be by recalling the mean and the variance of a binomial distribution and the relationship between the two distributions. A binomial distribution has mean np and variance $np(1 - p) = np - (np)p$. Now, if n gets large and p becomes small but $np = \lambda$ remains constant, the variance $np - (np)p = \lambda - \lambda p$ should tend toward λ. In fact, the Poisson distribution does have both its mean and its variance equal to λ.

The mean of the Poisson distribution can easily be derived formally if one remembers a simple Taylor series expansion of e^x—namely,

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

Then,

$$
\begin{aligned}
E(Y) &= \sum_y p(y) \\
&= \sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda} \\
&= \sum_{y=1}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} \\
&= \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \right) \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

The formal derivation of the fact that

$$V(Y) = \lambda$$

is left as a challenge for the interested reader. [Hint: First, find E[$Y(Y$ -1)].]

Example 4.21:

The manager of an industrial plant is planning to buy a new machine of either type A or type B. For each day's operation, the number of repairs X that machine A requires is a Poisson random variable with mean 0.10 t, where t denotes the time (in hours) of daily operation. The number of daily repairs Y for machine B is a Poisson random variable = with mean 0.12t. The daily cost of operating A is $C_A(t)$ =20t + 40X^2; for B, the cost is $C_B(t) = 16t + 40Y^2$. Assume that the repairs take negligible time and that each night the machines are to be cleaned, so that they operate like new machines at the start of each day. Which machine minimizes the expected daily cost if a Y consists of the following time spans?

1. 10 hours
2. 20 hours

Solution:

The expected cost for machine A is

$$
\begin{aligned}
E[C_A(t)] &= 20t + 40E\left(X^2\right) \\
&= 20t + 40\left[V(X) + (E(X))^2\right] \\
&= 20t + 40\left[0.1t + 0.01t^2\right] \\
&= 24t + 0.4t^2
\end{aligned}
$$

Similarly,

$$\begin{aligned}E[C_B(t)] &= 16t + 40E(X^2) \\ &= 16t + 40[V(X) + (E(X))^2] \\ &= 16t + 40[0.12t + 0.0144t^2] \\ &= 20.8t + 0.576t^2\end{aligned}$$

1. Here,

$$E[C_A(10)] = 24(10) + 0.4(10)^2 = 280$$

and

$$E[C_B(10)] = 20.8(10) + 0.576(10)^2 = 265.60$$

which results in the choice of machine *B*.
2. Here,

$$E[C_A(20)] = 24(20) + 0.4(20)^2 = 640$$

and

$$E[C_B(20)] = 20.8(20) + 0.576(20)^2 = 646.40$$

which results in the choice of machine *A*.

4.7.3. History and Applications

Siméon-Denis Poisson (1781-1840), examiner and professor at the *École Polytechnique* of Paris for nearly forty years, wrote over 300 papers in the fields of mathematics, physics and astronomy (see Figure 4.14). His most important works were a series of papers on definite integrals and his advances on Fourier series, providing a foundation for later work in this area by Dirichlet and Riemann. However, it was his derivation of the exponential limit of the binomial distribution, much as we saw above, for which he is best known in probability and statistics. The derivation was given no special emphasis, Cournot republished it in 1843 with calculations demonstrating the effectiveness of the approximation. Although De Moivre had presented the exponential limit of the binomial distribution in the first edition of *The Doctrine of Chances*, published in 1718, this distribution became known as the Poisson distribution.

The Poisson distribution was rediscovered by von Bortkiewicz in 1898. He tabulated the number of cavalrymen in the Prussian army who died from a kick from a horse during a year. To see how the Poisson distribution applies, first suppose that a cavalryman can either be killed by a horsekick during a year or not. Further suppose that the chance of this rare event is the same for all soldiers and that soldiers have independent chances of being killed. Thus, the number of cavalrymen killed during a year is a binomial random variable. However, the probability of being killed, *p*, is very small and the number of cavalrymen (trials) is very large. Therefore, the Poisson limit is a reasonable model for the data. A comparison of the observed and theoretical relative frequencies is shown in Table 4.1. Notice that the two agree well, indicating the Poisson is an adequate model for these data.

Figure 4.14. Siméon-Denis Poisson (1781-1840)

Source: http://en.wikipedia.org/wiki/Simeon_Poisson

Table 4.1. Deaths of Prussian Calvary Men Due to Kick by a Horse

Number of Calvalrymen Killed During a Year	Frequency	Relative Frequency	Theoretical Probability
0	109	0.545	0.544
1	65	0.325	0.331
2	22	0.110	0.191
3	3	0.015	0.021
4	1	0.005	0.003

The Poisson distribution can be used as an approximation for the binomial (large n and small p). When count data are observed, the Poisson model is often the first model considered. If the estimates of mean and variance differ significantly so that this property of the Poisson is not reasonable for a particular application, then one turns to other discrete distributions, such as the binomial (or negative binomial) for which the variance is less (greater) than the mean. The number of radioactive particles emitted in a given time period, number of telephone calls received in a given time period, number of equipment failures in a given time period, the number of defects in a specified length of wire, and the number of insects in a specified volume of soil are some of the many types of data that have been modeled using the Poisson distribution.

The Poisson Distribution

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \qquad y = 0,\ 1,\ 2,\ .. \qquad for\ \lambda > 0$$

$$E(Y) = \lambda \qquad\qquad V(Y) = \lambda$$

Exercises

4.64. Let Y denote a random variable that has a Poisson distribution with mean $\lambda = 4$. Find the following probabilities.

a. $P(Y = 5)$
b. $P(Y < 5)$
c. $P(Y \geq 5)$
d. $P(Y \geq 5 | Y \geq 2)$

4.65. The number of calls coming into a hotel's reservation center averages three per minute.
a. Find the probability that no calls will arrive in a given one-minute period.
b. Find the probability that at least two calls will arrive in a given one-minute period.
c. Find the probability that at least two calls will arrive in a given two-minute period.

4.66. The Meteorology Department of the University of Hawaii modeled the number of hurricanes coming within 250 nautical miles of Honolulu during a year using a Poisson distribution with a mean of 0.45 (http://www.math.hawaii.edu/~ramsey/Hurricane.html). Using this model, determine the probabilities of the following events.
a. At least one hurricane will come within 250 nautical miles of Honolulu during the next year
b. At most four hurricanes will come within 250 nautical miles of Honolulu during the next year

4.67. A certain type of copper wire has a mean number of 1.5 flaws per meter.
a. Justify using the Poisson distribution as a model for the number of flaws in a certain length of this wire.
b. Find the probability of having at least one flaw in a meter length of the copper wire.
c. Find the probability of having at least one flaw in a 5-meter length of the copper wire.

4.68. Referring to Exercise 4.67, the cost of repairing the flaw in the copper wire is $8 per flaw. Find the mean and the standard deviation of the repair costs for a 10-meter length of wire in question.

4.69. Customer arrivals at a checkout counter in a department store have a Poisson distribution with an average of seven per hour. For a given hour, find the probabilities of the following events
a. Exactly seven customers arrive
b. No more than two customers arrive
c. At least two customers arrive.

4.70. Referring to Exercise 4.69, if it takes approximately 10 minutes to service each customer, find the mean and the variance of the total service time connected to the customer arrivals for one hour. (Assume that an unlimited number of servers are available, so that no customer has to wait for service.) Is total service time highly likely to exceed 200 minutes?

4.71. Referring to Exercise 4.69, find the probabilities that exactly two customers will arrive in the follow two-hour periods of time.
a. Between 2:00 p.m. and 4:00 p.m. (one continuous two-hour period).

b. Between 1:00 p.m. and 2:00 p.m. and between 3:00 p.m. and 4:00 p.m.(two separate one-hour periods for a total of two hours).

4.72. The number of grasshoppers per square meter of rangeland is often well modeled using the Poisson distribution. Suppose the mean number of grasshoppers in a specified region that has been grazed by cattle is 0.5 grasshoppers per square meter. Find the probabilities of the following events.
a. Five or more grasshoppers in a randomly selected square meter in this region
b. No grasshoppers in a randomly selected square meter in this region
c. At least one grasshopper in a randomly selected square meter in this region

4.73. The number of particles emitted by a radioactive source is generally well modeled by the Poisson distribution. If the average number of particles emitted by the source in an hour is four, find the following probabilities.
a. The number of emitted particles in a given hour is at least 6
b. The number of emitted particles in a given hour will be at most 3
c. No particles will be emitted in a given 24-hour period

4.74. Chu (2003) studied the number of goals scored during the 232 World Cup soccer games played from 1990 to 2002. Only goals scored during the 90 minutes of regulation play were considered. The average number of goals scored each game was 2.5. Assuming this mean continues to hold for other World Cup games, find the probabilities associated with the following events.
a. At least six goals are scored during the 90 minutes of regulation play in a World Cup game
b. No goals are scored during the 90 minutes of regulations play in a World Cup game

4.75. When developing a voice network, one important consideration is "the availability of service." One measure of this is "congestion." Congestion is the probability that the next call will be blocked. If there are c circuits, then congestion is the probability that c or more calls are in progress. A large sales firm has an average of 8 calls during any minute of the business day.
a. If the firm has 10 circuits, what is the probability that there will be congestion during any minute of the business day?
b. The firm has decided to upgrade its system and wants to add enough circuits so that the probability of congestion is no more than 2% during any minute of the business day. How many circuits does the firm need?
c. If the firm wants to add enough circuits that it could handle a 20% growth in telephone traffic and still have a probability of congestion of no more than 2% during any minute of the business day, how many circuits should it have?

4.76. The number of fatalities due to shark attack during a year is modeled using a Poisson distribution. The International Shark Attack File (ISAF) investigates shark-human interactions worldwide. Internationally, an average of 4.4 fatalities per year occurred during 2001 to 2005. Assuming that this mean will remain constant for the next five years (2006 to 2010), find the probabilities of the following events.

a. No shark fatalities will be recorded in a given year.
b. Sharks will cause at least 6 human deaths in a given year
c. No shark fatalities will occur in 2006 to 2010
d. At most 12 shark fatalities will occur in 2006 to 2010

4.77. Schmuland (2001) explored the use of the Poisson distribution to model the number of goals the hockey star, Wayne Gretzky, scored during a game as an Edmonton Oiler. Gretzky played 696 games with the following distribution of the number of goals scored:

Points	0	1	2	3	4	5	6	7	8	9
Number of Games	69	155	171	143	59	77	14	6	2	0

a. Find the average number of goals scored per game.
b. Using the average found in (a) and assuming that Gretzky's goals were scored according to a Poisson distribution, find the expected number of games in which 0, 1, 2, …, ≥ 9 goals were scored.
c. Compare the actual numbers of games with the expected numbers found in (b). Does the Poisson seem to be a reasonable model?

4.78. The number of bacteria colonies of a certain type in samples of polluted water has a Poisson distribution with a mean of two per cubic centimeter.
a. If four 1-cubic-centimeter samples of this water are independently selected, find the probability that at least one sample will contain one or more bacteria columns.
b. How many 1-cubic-centimeter samples should be selected to establish a probability of approximately 0.95 of containing at lest one bacteria colony?

4.79. Let Y have a Poisson distribution with mean λ. Find $E[Y(Y-1)]$, and use the result to show that $V(Y) = \lambda$.

4.80. A food manufacturer uses an extruder (a machine that produces bite-size foods, like cookies and many snack foods) that has a revenue-producing value to the firm of \$300 per hour when it is in operation. However, the extruder breaks down an average of twice every 10 hours of operation. If Y denotes the number of breakdowns during the time of operation, the revenue generated by the machine is given by

$$R = 300t - 75Y^2$$

where t denotes hours of operations. The extruder is shut down for routine maintenance on a regular schedule, and it operates like a new machine after this maintenance. Find the optimal maintenance interval t_0 to maximize the expected revenue between shutdowns.

4.8. The Hypergeometric Distribution

4.8.1 The Probability Function

The distributions already discussed in this chapter have as their basic building block a series of *independent* Bernoulli trials. The examples, such as sampling from large lots, depict situations in which the trials of the experiment generate, for all practical purposes, independent outcomes. But suppose that we have a relatively small lot consisting of N items, of which k are defective. If two items are sampled sequentially, the outcomes for the second draw is significantly influenced by what happened on the first draw, provided that the first item drawn remains out of the lot. A new distribution must be developed to handle this situation involving *dependent* trials.

In general, suppose that a lot consists of N items, of which k are of one type (called *successes*) and $N - k$ are of another type (called *failures*). Suppose that n items are sampled randomly and sequentially from the lot, and suppose that none of the sampled items is replaced. (This is called *sampling without replacement*.) Let $X_i = 1$ if the ith draw results in a success, and let $X_i = 0$ otherwise, where $i = 1, 2, \ldots, n$. Let Y denote the total number of successes among the n sampled items. To develop the probability distribution for Y, let us start by looking at a special case for $Y = y$. One way for successes to occur is to have

$$X_1 = 1, \qquad X_2 = 1, \qquad \ldots, \qquad X_y = 1, \qquad X_{y+1} = 0, \qquad \ldots, \qquad X_n = 0$$

We know that

$$P(X_1 = 1, X_2 = 1) = P(X_1 = 1)P(X_2 = 1 \mid X_1 = 1)$$

and this result can be extended to give

$$\begin{aligned} &P(X_1 = 1, X_2 = 1, \ \ldots, \ X_y = 1, X_{y+1} = 0, \ \ldots, X_n = 0) \\ &\quad = P(X_1 = 1)P(X_2 = 1 \mid X_1 = 1)P(X_3 = 1 \mid X_2 = 1, X_1 = 1)\cdots \\ &\qquad P(X_n = 0 \mid X_{n-1} = 0, \ \ldots, X_{y+1} = 0, X_y = 1, \ \ldots, X_1 = 1) \end{aligned}$$

Now,

$$P(X_1 = 1) = \frac{k}{N}$$

if the item is randomly selected; and similarly,

$$P(X_2 = 1 \mid X_1 = 1) = \frac{k-1}{N-1}$$

because, at this point, one of the k successes has been removed. Using this idea repeatedly, we see that

$$P(X_1 = 1, X_2 = 1, \ \ldots, \ X_y = 1, X_{y+1} = 0, \ \ldots, X_n = 0)$$
$$= \left(\frac{k}{N}\right)\left(\frac{k-1}{N-1}\right)\cdots\left(\frac{k-y+1}{N-y+1}\right)\times\left(\frac{N-k}{N-y}\right)\cdots\left(\frac{N-k-n+y+1}{N-n+1}\right)$$

provided that $y \le k$. A more compact way to write the preceding expression is to employ factorials, arriving at the formula

$$\frac{\dfrac{k!}{(k-y)!}\times\dfrac{(N-k)!}{(N-k-n+y)!}}{\dfrac{N!}{(N-n)!}}$$

(The reader can check the equivalence of the two expressions.)

Any specified arrangement of y successes and $(n - y)$ failures will have the same probability as the one just derived for all successes followed by all failures; the terms will merely be rearranged. Thus, to find $P(Y = y)$, we need only count how many of these arrangements are possible. Just as in the binomial case, the number of such arrangements is $\binom{n}{y}$. Hence, we have

$$P(Y = y) = \binom{n}{y}\frac{\dfrac{k!}{(k-y)!}\times\dfrac{(N-k)!}{(N-k-n+y)!}}{\dfrac{N!}{(N-n)!}}$$
$$= \frac{\binom{k}{y}\binom{N-k}{n-y}}{\binom{N}{n}}$$

Of course, $0 \le y \le k \le N$ and $0 \le y \le n \le N$. This formula is referred to as the *hypergeometric probability distribution*. Notice that it arises from a situation quite similar to the binomial, except that the trials here are *dependent*.

Finding the probabilities associated with the hypergeometric distribution can be computationally intensive, especially as N and n increase. Calculators and computer software are often valuable tools in determining hypergeometric probabilities. Explore the shape of the hypergeometric distribution further using the applet.

Example 4.22:
Two positions are open in a company. Ten men and five women have applied for a job at this company, and all are equally qualified for either position. The manager randomly hires 2 people from the applicant pool to fill the positions. What is the probability that a man and a woman were chosen.

Solution:

If the selections are made at random, and if Y denotes the number of men selected, then the hypergeometric distribution would provide a good model for the behavior of Y. Hence,

$$P(Y=1)=p(1)=\frac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}}=\frac{(10)(5)}{\left(\frac{15(14)}{2}\right)}=\frac{10}{21}$$

Here, $N = 15$, $k = 1$, $n = 2$, and $y = 1$.

3.8.2. Mean and Variance

The techniques needed to derive the mean and the variance of the hypergeometric distribution will be given in Chapter 6. The results are

$$E(Y)=n\left(\frac{k}{N}\right)$$
$$V(Y)=n\left(\frac{k}{N}\right)\left(1-\frac{k}{n}\right)\left(\frac{N-n}{N-1}\right)$$

Because the probability of selecting a success on one draw is k/n, the mean of the hypergeometric distribution has the same form as the mean of the binomial distribution. Likewise, the variance of the hypergeometric matches the variance of the binomial, multiplied by $(N-n)/(N-1)$, a correction factor for dependent samples.

Example 4.23:

In an assembly-line production of industrial robots, gear-box assemblies can be installed in one minute each if the holes have been properly drilled in the boxes, and in 10 minutes each if the holes must be redrilled. Twenty gear boxes are in stock, and two of these have improperly drilled holes. Five gear boxes are selected from the twenty available for installation in the next five robots in line.
1. Find the probability that all five gear boxes will fit properly.
2. Find the expected value, the variance, and the standard deviation of the time it will take to install these five gear boxes.

Solution:

In this problem, $N = 20$; and the number of nonconforming boxes is $k = 2$, according to the manufacturer's usual standards. Let Y denote the number of nonconforming boxes (that is, the number with improperly drilled holes) in the sample of five. Then,

$$P(Y=0)=\frac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}}$$

$$=\frac{(1)(8568)}{15{,}504}$$

$$=0.55$$

The total time *T* taken to install the boxes (in minutes) is

$$T=10Y+(5-Y)$$

$$=9Y+5$$

since each of *Y* nonconforming boxes takes 10 minutes to install, and the others take only one minute. To find $E(T)$ and $V(T)$, we first need to calculate $E(Y)$ and $V(Y)$:

$$E(Y)=n\left(\frac{k}{N}\right)=5\left(\frac{2}{20}\right)=0.5$$

and

$$V(Y)=n\left(\frac{k}{n}\right)\left(1-\frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$=5(0.1)(1-0.1)\left(\frac{20-5}{20-1}\right)$$

$$=0.355$$

It follows that

$$E(T)=9E(Y)+5$$

$$=9(0.5)+5$$

$$=9.5$$

and

$$V(T)=(9)^2V(Y)$$

$$=81(0.355)$$

$$=28.755$$

Thus, installation time should average 9.5 minutes, with a standard deviation of $\sqrt{28.755}=5.4$ minutes.

4.8.3. History and Applications

Although they did not use the term hypergeometric distribution, Bernoulli and de Moivre used the distribution to solve some of the probability problems they encountered. In 1899, Karl Pearson (see Figure 4.15) discussed using the "hypergeometrical series)" to model data. But, it was not until 1936 that the term "hypergeometric distribution" actually appeared in the literature (http://members.aol.com/jeff570/h.html).

Figure 4.15. Karl Pearson

Source: http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Pearson.html

The primary application of the hypergeometric distribution is in the study of finite populations. Although most populations are finite, many are large enough so that the probabilities are relatively stable as units are drawn. However, as the fraction of the population sampled becomes larger, the probabilities begin to change significantly with each new unit selected for the sample. The hypergeometric distribution has been used extensively in discrimination cases, quality control, and surveys.

The Hypergeometric Distribution

$$P(Y=y)=\frac{\binom{k}{y}\binom{N-k}{n-y}}{\binom{N}{n}}, \qquad y=0,\ 1,\ \ldots,k \ \textit{with} \ \binom{b}{a}=0 \ \textit{if} \ a>b$$

$$E(Y)=n\left(\frac{k}{N}\right) \qquad \text{and} \qquad V(Y)=n\left(\frac{k}{N}\right)\left(1-\frac{k}{n}\right)\left(\frac{N-n}{N-1}\right)$$

Exercises

4.81. From a box containing five white and four red balls, two balls are selected at random, without replacement. Find the probabilities of the following events.
a. Exactly one white ball is selected
b. At least one white ball is selected
c. Two white balls are selected, given that at lest one white ball is selected
d. The second ball drawn is white

4.82. A company has ten personal computers (PCs) in its warehouse. Although all are new and still in boxes, four do not currently function properly. One of the company's offices requests five PCs, and the warehouse foreman selects five from the stock of ten and ships to the requesting office. What is the probability that all five of the PC are not defective?

4.83. Referring to Exercise 4.82, the office requesting the personal computers returns the defective ones for repair. If it costs \$80 to repair each PC, find the mean and the variance of the total repair cost. By Tchebysheff's Theorem, in what interval should you expect, with a probability of at least 0.95, the repair costs of these five PCs to lie?

4.84. A foreman has ten employees, and the company has just told him that he must terminate four of them. Three of the ten employees belong to a minority ethnic group. The foreman selected all three minority employees (plus one other) to terminate. The minority employees then protested to the union steward that they were discriminated against by the foreman. The foreman claimed that the selection had been completely random. What do you think? Justify your anwer.

4.85. A small firm has 25 employees. Eight are single, and the other 17 are married. The owner is contemplating a change in insurance coverage, and she randomly selects five people to get their opinions.
a. What is the probability that she only talks to single people?
b. What is the probability that she only talks to married people?
c. What is the probability that 2 single and 3 married people are in the sample?

4.86. An auditor checking the accounting practices of a firm samples four accounts from an accounts receivable list of 12. Find the probability that the auditor sees at least one past-due account under the following conditions.
a. There are two such accounts among the twelve.
b. There are six such accounts among the twelve.
c. There are eight such accounts among the twelve.

4.87. The pool of qualified jurors called for a high profile case has 12 whites, 9 blacks, 4 Hispanics, and 2 Asians. From these, 12 will be selected to serve on the jury. Assume that all qualified jurors meet the criteria for serving. Find the probabilities of the following events.
a. No white is on the jury
b. All nine blacks serve on the jury
c. No Hispanics or Asians serve on the jury.

4.88. A student has a drawer with 20 AA batteries. However, the student does not realize that 3 of the batteries have lost their charge (will not work). She realizes that the batteries on her calculator are no longer working, and she is in a hurry to get to a test. She grabs two batteries from the drawer at random to replace the two batteries in her calculator. What is the probability that she will get two good batteries so that her calculator will work during the test?

4.89. A corporation has a pool of six firms (four of which are local) from which they can purchase certain supplies. If three firms are randomly selected without replacement, find the probabilities of the following events
a. All three selected firms are local
b. At least one selected firm is not local

4.90. Specifications call for a type of thermistor to test out at between 9000 and 1000 ohms at 25°C. Ten thermistors are available, and three of these are to be selected for use. Let Y denote the number among the three that do not conform to specifications. Find the probability distribution for Y (in tabular form) if the following conditions prevail.
a. The ten contain two thermistors not conforming to specifications
b. The ten contain four thermistors not conforming to specifications.

4.91. A group of 4 men and 6 women are faced with an extremely difficult and unpleasant task, requiring 3 people. They decide to draw straws to see who must do the job. Eight long and two short straws are made. A person outside the group is asked to hold the straws, and each person in the group selects one. Find the probability of the following events.
a. Both short straws are drawn by men
b. Both short straws are drawn by women
c. The short straws are drawn by a man and a woman

4.92. An eight-cylinder automobile engine has two misfiring spark plugs. If all four plugs are removed from one side of the engine, what is the probability that the two misfiring plugs are among them?

4.93. Lot acceptance sampling procedures for an electronics manufacturing firm call for sampling n items from a lot of N items and accepting the lot if $Y \leq c$, where Y is the number of nonconforming items in the sample. For an incoming lot of 20 transistors, 5 are to be sampled. Find the probability of accepting the lot if $c = 1$ and the actual number of nonconforming transistors in the lot are as follows
a. 0
b. 1
c. 2
d. 3

4.94. In the setting and terminology of exercise 4.93, answer the same questions if $c = 2$.

4.95. Two assembly lines (I and II) have the same rate of defectives in their production of voltage regulators. Five regulators are sampled from each line and tested. Among the total of ten tested regulators, four are defective. Find the probability that exactly two of the defectives came from line I.

4.96. A 10-acre area has N raccoons. Ten of these raccoons were captured, marked so they could be recognized, and released. After five days, twenty raccoons were captured. Let X denote the number of those captured on the second occasion that was marked during the first sampling occasion. Suppose that captures at both time points can be treated as random selections from the population and that the same N raccoons were in the area on both sampling occasions (no additions or deletions).
If $N = 30$, what is the probability that no more than 5 of those captured during the second sampling period were marked during the first sampling occasion?
If 12 raccoons in the second sample were marked from having been caught in the first one, what value of N would result in the probability of this happing being the largest?

4.9 The Moment-generating Function

We saw in earlier sections that, if $g(Y)$ is a function of a random variable Y, with probability distribution given by $p(y)$, then

$$E[g(Y)] = \sum_y g(y)p(y)$$

A special function with many theoretical uses in probability theory is the expected value of e^{tY}, for a random variable Y, and this expected value is called the moment-generating function (mgf). The definition of a moment generating function is given in Definition 4.8.

Definition 4.8. The **moment generating function** (mgf) of a random variable is denoted by $M(t)$ and defined to be

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

The expected values of powers of a random variable are often called moments. Thus, $E(Y)$ is the first moment of Y, and $E(Y^2)$ is the second moment of Y. One use for the moment-generating functions is that, in fact, it does generate moments of Y. When $M(t)$ exists, it is differentiable in a neighborhood of the origin $t = 0$, and the derivatives may be taken inside the expectation. Thus

$$\begin{aligned} M^{(1)}(t) &= \frac{dM(t)}{dt} \\ &= \frac{d}{dt}E(e^{tY}) \\ &= E\left[\frac{d}{dt}e^{tY}\right] \\ &= E\left[Ye^{tY}\right] \end{aligned}$$

Now, if we set $t = 0$, we have

$$M^{(1)}(0) = E(Y)$$

Going on to the second derivative,

$$M^{(2)}(t) = E(Y^2 e^{tY})$$

and

$$M^{(2)}(0) = E(Y^2)$$

In general,

$$M^{(k)}(0) = E(Y^k)$$

It often is easier to evaluate $M(t)$ and its derivatives than to find the moments of the random variable directly. Other theoretical uses of the mgf will be discussed in later chapters.

Example 4.24:

Evaluate the moment generating function for the geometric distribution, and use it to find the mean and the variance of this distribution.

Solution:

For the geometric variable Y, we have

$$
\begin{aligned}
M(t) &= E(e^{tY}) \\
&= \sum_{y=0}^{\infty} e^{ty} pq^{y} \\
&= p\sum_{y=0}^{\infty} \left(qe^{t}\right)^{y} \\
&= p\left[1 + qe^{t} + \left(qe^{t}\right)^{2} + \cdots\right] \\
&= p\left(\frac{1}{1-qe^{t}}\right) \\
&= \frac{p}{1-qe^{t}}
\end{aligned}
$$

because the series is geometric with a common ratio of qe^t. Note: For the series to converge and the moment generating function to exist, we must have $qe^t < 1$, which is the case if $t < \ln(1/q)$. It is important for the moment generating function to exist for t in a neighborhood about 0, and it does here.

To evaluate the mean, we have

$$
\begin{aligned}
M^{(1)}(t) &= \frac{0 + pqe^{t}}{\left(1-qe^{t}\right)^{2}} \\
&= \frac{pqe^{t}}{\left(1-qe^{t}\right)^{2}}
\end{aligned}
$$

and

$$M^{(1)}(t) = \frac{pq}{\left(1-q\right)^{2}} = \frac{q}{p}$$

To evaluate the variance, we first need

$$E\left(Y^{2}\right) = M^{(2)}(t)$$

Now,

$$
\begin{aligned}
M^{(2)}(t) &= \frac{\left(1-qe^{t}\right)^{2} pqe^{t} - pqe^{t}(2)\left(1-qe^{t}\right)(-1)qe^{t}}{\left(1-qe^{t}\right)^{4}} \\
&= \frac{pqe^{t} + pq^{2}e^{2t}}{\left(1-qe^{t}\right)^{3}}
\end{aligned}
$$

and

$$M^{(2)}(0) = \frac{pq + pq^{2}}{\left(1-q\right)^{3}} = \frac{q(1+q)}{p^{2}}$$

Hence,

$$V(Y) = E\left(Y^2\right) - [E(Y)]^2$$
$$= \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2$$
$$= \frac{q}{p^2}$$

Moment-generating functions have important properties that make them extremely useful in finding expected values and in determining the probability distributions of random variables. These properties will be discussed in detail in Chapters 5 and 6, but one such property is given in Exercise 4.109.

4.10. The Probability-generating Function

In an important class of discrete random variables, Y takes integral values ($Y = 0, 1, 2, 3, \ldots$) and, consequently, represents a count. The binomial, geometric, hypergeometric, and Poisson random variables all fall in this class. The following examples present practical situations involving integral-valued random variables. One, tied to the theory of queues (waiting lines), is concerned with the number of persons (or objects) awaiting service at a particular point in time. Understanding the behavior of this random variable is important in designing manufacturing plants where production consists of a sequence of operations, each of which requires a different length of time to complete. An insufficient number of service stations for a particular production operation can result in a bottleneck—the forming of a queue of products waiting to be serviced—which slows down the entire manufacturing operation. Queuing theory is also important in determining the number of checkout counters needed for a supermarket and in designing hospitals and clinics.

Integer-valued random variables are extremely important in studies of population growth, too. For example, epidemiologists are interested in the growth of bacterial populations and also in the growth of the number of persons afflicted by a particular disease. The number of elements in each of these populations is an integral-valued random variable.

A mathematical device that is very useful in finding the probability distributions and other properties of integral-valued random variables in the probability-generating function $P(t)$, which is defined in Definition 4.9.

Definition 4.9. The **probability generating function** of a random variable is denoted by $P(t)$ and is defined to be

$$P(t) = E(t^Y)$$

If Y is an integer-valued random variable, with

$$P(Y = y) = p_i, \qquad y = 0,\ 1,\ 2,\ \ldots$$

Then

$$P(t) = E(t^Y) = p_0 + p_1 t + p_2 t^2 + \ldots$$

The reason for calling $P(t)$ a probability–generating function is clear when we compare $P(t)$ with the moment-generating function $M(t)$. Particularly, the coefficient of t^i in $P(t)$ is the probability p_i. If we know $P(t)$ and can expand it into a series, we can determine $p(y)$ as the coefficient of t^y. Repeated differentiation of $P(t)$ yields *factorial moments* for the random variable Y

> Definition 4.7: The kth **factorial moment** for a random variable Y is defined to be
>
> $$\mu_{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-k+1)]$$
>
> where k is a positive integer.

When a probability-generating function exists, it can be differentiated in a neighborhood of $t = 1$. Thus, with

$$P(t) = E(t^Y)$$

we have

$$\begin{aligned} P^{(1)}(t) &= \frac{dP(t)}{dt} \\ &= \frac{d}{dt}E(t^Y) \\ &= E\left[\frac{d}{dt}t^Y\right] \\ &= E\left[Yt^{Y-1}\right] \end{aligned}$$

Setting $t = 1$, we have

$$P^{(1)}(1) = E(Y)$$

Similarly,

$$P^{(2)}(t) = E[Y(Y-1)t^{Y-2}]$$

and

$$P^{(2)}(1) = E[Y(Y-1)]$$

In general,

$$P^{(k)}(t) = E[Y(Y-1)\cdots(Y-k+1)t^{Y-k}]$$

and

$$\begin{aligned} P^{(k)}(1) &= E[Y(Y-1)\cdots(Y-k+1)] \\ &= \mu_{[k]} \end{aligned}$$

Example 4.25:

Find the probability-generating function for the geometric random variable, and use this function to find the mean.

Solution:

$$P(t) = E\left(t^Y\right) = \sum_{y=0}^{\infty} t^y pq^y$$

$$= p\sum_{y=0}^{\infty} (qt)^y$$

$$= \frac{p}{1-qt}$$

where qt < 1 for the series to converge. Now,

$$P^{(1)}(t) = \frac{d}{dt}\left(\frac{p}{1-qt}\right) = \frac{pq}{(1-qt)^2}$$

Setting $t = 1$,

$$P^{(1)}(1) = \frac{pq}{(1-q)^2} = \frac{q}{p}$$

which is the mean of a geometric random variable.

Since we already have the moment-generating function to assist us in finding the moment of a random variable, we might ask how knowing $P(t)$ can help us. The answer is that in some instances it may be exceedingly difficult to find $M(t)$ but easy to find $P(t)$. Alternatively, $P(t)$ may be easier to work with in a particular setting. Thus, $P(t)$ simply provides an additional tool for finding the moments of a random variable. It may or may not be useful in a given situation.

Finding the moments of a random variable is not the major use of the probability-generating function. Its primary application is in deriving the probability function (and hence the probability distribution) for related integral-valued random variables. For these applications, see Feller (1968), Parzen (1964), and Section 7.7.

Exercises

4.97. Find the moment-generating function for the Bernoulli random variable.

4.98. Derive the mean and variance of the Bernoulli random variable using the moment generating function derived in Exercise 4.97.

4.99. Show that the moment-generating function for the binomial random variable is given by

$$M(t) = \left(pe^t + q\right)^n$$

4.100. Derive the mean and variance of the binomial random variable using the moment-generating function derived in Exercise 4.99.

4.101. Show that the moment-generating function for the Poisson random variable with mean λ is given by

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

4.102. Derive the mean and variance of the Poisson random variable using the moment-generating function derived in Exercise 4.101.

4.103. Show that the moment-generating function for the negative binomial random variable is given by

$$M(t) = \left(\frac{p}{1 - qe^t}\right)^r$$

4.104. Derive the mean and variance of the negative binomial random variable using the moment-generating function derived in Exercise 4.103.

4.105. Derive the probability-generating function of a Poisson random variable with parameter λ.

4.106. Using the probability-generating function derived in Exercise 4.105 find the first and second factorial moments of a Poisson random variable. From the first two factorial moments, find the mean and variance of the Poisson random variable.

4.107. Derive the probability-generating function of the binomial random variable of n trials, with probability of success p.

4.108. Find the first and second factorial moments of the binomial random variable in Exercise 4.107. Using the first two factorial moments, find the mean and variance of the Poisson random

4.109. If X is a random variable with moment-generating function $M(t)$, and Y is a function of X given by $Y = aX + b$, show that the moment-generating function for Y is $e^{tb}M(at)$.

4.110. Use the result of Exercise 4.109 to show that

$$E(Y) = aE(X) + b$$

and

$$V(Y) = a^2 V(X)$$

4.11 Markov Chains

Consider a system that can be in any of a finite number of states. Assume that the system moves from state to state according to some prescribed probability law. The system, for example, could record weather conditions from day to day, with the possible states being clear, partly cloudy, and cloudy. Observing conditions over a long period of time would allow one to find the probability of its being clear tomorrow given that it is partly cloudy today.

Let X_i denote the state of the system at time point i, and let the possible states be denoted by $S_1, \ldots, S_m$, for a finite integer m. We are interested not in the elapsed time between transitions from one time state to another, but only in the states and the probabilities of going from one state to another—that is, in the *transition probabilities*. We assume that

$$P(X_i = S_k \mid X_{i-1} = S_j) = p_{jk}$$

where p_{jk} is the transition probability from S_j to S_k; and this probability is independent of i. Thus, the transition probabilities depend not on the time points, but only on the states. The event $(X_i = S_k | X_{i-1} = S_j)$ is assumed to be independent of the past history of the process. Such a process is called a Markov chain with stationary transition probabilities. The transition probabilities can conveniently be displayed in a matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Let X_0 denote the starting state of the system, with probabilities given by

$$P_k^{(0)} = P(X_0 = S_k)$$

and let the probability of being in state S_k after n steps be given by $p_k^{(n)}$. These probabilities are conveniently displayed by vectors:

$$\mathbf{p}^{(0)} = [p_1^{(0)}, p_2^{(0)}, \ldots, p_m^{(0)}]$$

and

$$\mathbf{p}^{(n)} = [p_1^{(n)}, p_2^{(n)}, \ldots, p_m^{(n)}]$$

To see how $\mathbf{p}^{(0)}$ and $\mathbf{p}^{(1)}$ are related, consider a Markov chain with only two states, so that

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

There are two ways to get to state 1 after one step: either the chain starts in state 1 and stays there, or the chain starts in state 2 and then moves to state 1 in one step. Thus,

$$\mathbf{p}_1^{(1)} = p_1^{(0)} p_{11} + p_2^{(0)} p_{21}$$

Similarly,

$$\mathbf{p}_2^{(1)} = p_1^{(0)} p_{12} + p_2^{(0)} p_{22}$$

In terms of matrix multiplication,

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)}\mathbf{P}$$

and, in general,

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)}\mathbf{P}$$

P is said to be regular if some power of **P** ($\mathbf{P}^n$ for some n) has all positive entries. Thus, one can get from state S_j to state S_k, eventually, for any pair (j, k). (Notice that the condition of regularity rules out certain chains that periodically return to certain states.) If **P** is regular, the chain has a stationary (or equilibrium) distribution that gives the probabilities of its being in the respective states after many transitions have evolved. In other words, $p_j^{(n)}$ must have a limit π_j, as $n \to \infty$. Suppose that such limits exist; then $\pi = (\pi_1, \ldots, \pi_n)$ must satisfy

$$\pi = \pi\mathbf{P}$$

because $\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)}\mathbf{P}$ will have the same limit as $\mathbf{p}^{(n-1)}$.

Example 4.26:

A supermarket stocks three brands of coffee—A, B, and C—and customers switch from brand to brand according to the transition matrix

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 2/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

where S_1 corresponds to a purchase of brand A, S_2 to brand B, and S_3 to brand C; that is, ¾ of the customers buying brand A also buy brand A the next time they purchase coffee, whereas ¼ of these customers switch to brand B.

1. Find the probability that a customer who buys brand A today will again purchase brand A two weeks from today, assuming that he or she purchases coffee once a week
2. In the long run, what fractions of customers purchase the respective brands?

Solution:

1. Assuming that the customer is chosen at random, his or her transition probabilities are given by **P**. The given information indicates that $\mathbf{p}^{(0)} = (1, 0, 0)$; that is, the customer starts with a purchase of brand A. Then

$$p^{(1)} = p^{(0)}P = \left(\frac{3}{4}, \frac{1}{4}, 0\right)$$

gives the probabilities for the next week's purchase. The probabilities for two weeks from now are given by

$$p^{(2)} = p^{(1)}P = \left(\frac{9}{16}, \frac{17}{48}, \frac{1}{12}\right)$$

That is, the chance of the customer's purchasing A two weeks from now is only 9/16.
2. The answer to the long-run frequency ratio is given by π, the stationary distribution. The equation

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$$

yields the system of equations

$$\pi_1 = \left(\frac{3}{4}\right)\pi_1 + \left(\frac{1}{4}\right)\pi_3$$

$$\pi_2 = \left(\frac{1}{4}\right)\pi_1 + \left(\frac{2}{3}\right)\pi_2 + \left(\frac{1}{4}\right)\pi_3$$

$$\pi_3 = \left(\frac{1}{3}\right)\pi_2 + \left(\frac{1}{2}\right)\pi_3$$

Combining these equations with the fact that $\pi_1 + \pi_2 + \pi_3 = 1$ yields

$$\pi = \left(\frac{2}{7}, \frac{3}{7}, \frac{2}{7}\right)$$

Thus, the store should stock more brand B coffee than either brand A or brand C.

Example 4.27:

Markov chains are used in the study of probabilities connected to genetic models. Recall from Section 3.2 that genes come in pairs. For any trait governed by a pair of genes, an individual may have genes that are homozygous dominant (GG), heterozygous (Gg), or homozygous recessive (gg). Each offspring inherits one gene of a pair from each parent, at random and independently.

Suppose that an individual of unknown genetic makeup is mated with a heterozygous individual. Set up a transition matrix to describe the possible states of a resulting offspring and their probabilities. What will happen to the genetic makeup of the offspring after many generations of mating with a heterozygous individual?

Solution:

If the unknown is homozygous dominant (GG) and is mated with a heterozygous individual (Gg), the offspring has a probability of ½ of being homozygous dominant and a probability of ½ of being heterozygous. If two heterozygous individuals are mated, the offspring may be homozygous dominant, heterozygous, or homozygous recessive with probabilities ¼, ½, and ¼, respectively. If the unknown is homozygous recessive (gg), the offspring of it and a heterozygous individual has a probability of ½ of being homozygous recessive and a probability of ½ of being heterozygous. Following along these lines, a transition matrix from the unknown parent to an offspring is given by

$$P = \begin{array}{c} \\ d \\ h \\ r \end{array} \begin{array}{c} \begin{array}{ccc} d & h & r \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

The matrix $\mathbf{P}^2$ has all positive entries (**P** is regular); and hence a stationary distribution exists. From the matrix equation

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$$

we obtain

$$\begin{aligned} \pi_1 &= \left(\frac{1}{2}\right)\pi_1 + \left(\frac{1}{2}\right)\pi_2 \\ \pi_2 &= \left(\frac{1}{2}\right)\pi_1 + \left(\frac{1}{2}\right)\pi_2 + \left(\frac{1}{2}\right)\pi_3 \\ \pi_3 &= \left(\frac{1}{2}\right)\pi_2 \qquad\qquad + \left(\frac{1}{2}\right)\pi_3 \end{aligned}$$

Since $\pi_1 + \pi_2 + \pi_3 = 1$, the second equation yields $\pi_2 = \frac{1}{2}$. It is then easy to establish that $\pi_1 = \pi_3 = \frac{1}{4}$. Thus,

$$\pi = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$$

No matter what the genetic makeup of the unknown parent happened to be, the ratio of homozygous dominant to heterozygous to homozygous recessive offspring among its descendants, after many generations of mating with heterozygous individuals, should be 1:2:1.

An interesting example of a transition matrix that is not regular is formed by a Markov chain with absorbing states. A state S_i is said to be absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $j \neq i$. That is, once the system is in state S_i, it cannot leave it. The transition matrix for such a chain can always be arranged in a standard form, with the absorbing states listed first. For example, suppose that a chain has five states, of which two are obsorbing. Then **P** can be written as

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$$

where **I** is a 2 x 2 identity matrix and **0** is a matrix of zeros. Such a transition matrix is not regular. Many interesting properties of these chains can be expressed in terms of **R** and **Q** (see Kemeny and Snell, 1983).

The following discussion will be restricted to the case in which **R** and **Q** are such that it is possible to get to an absorbing state from every other state, eventually. In that case, the Markov chain eventually will end up in an absorbing state. Questions of interest then involve the expected number of steps to absorption and the probability of absorption in the various absorbing states.

Let m_{ij} denote the expected (or mean) number of times the system is in state S_j, given that it started in S_i, for nonabsorbing states S_i and S_j. From S_i, the system could go to an absorbing state in one step, or it could go to a nonabsorbing state—say S_k—and eventually be absorbed from there. Thus, m_{ij} must satisfy

$$m_{ij} = \partial_{ij} + \sum_k p_{ik} m_{kj}$$

Where the summation is over all nonabsorbing states and

$$\partial_{ij} = \begin{cases} 1, & if\ i = j \\ 0, & otherwise \end{cases}$$

The term ∂_{ij} accounts for the fact that, if the chain goes to an absorbing state in one step, it was in state S_i one time.

If we denote the matrix of m_{ij} terms by **M**, the preceding equation can then be generalized to

$$\mathbf{M} = \mathbf{I} + \mathbf{QM}$$

or

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}$$

Matrix operations, such as inversion, will not be discussed here. The equations can be solved directly if matrix operations are unfamiliar to the reader.)

The expected number of steps to absorption, from the nonabsorbing starting state S_i, will be denoted by $\mathbf{m}_i$ and given simply by

$$m_i = \sum_k m_{ik}$$

again summing over nonabsorbing states.

Turning now to the probability of absorption into the various absorbing states, we let a_{ij} denote the probability of the system's being absorbed in state S_j, given that it started in state S_i, for nonabsorbing S_i and *absorbing* S_j. Repeating the preceding argument, the system could move to S_j in one step, or it could move to a nonabsorbing state S_k and be absorbed from there. Thus, a_{ij} satisfies

$$a_{ij} = p_{ij} + \sum_k p_{ik} a_{kj}$$

where the summation occurs over the nonabsorbing states. If we denote the matrix of a_{ij} terms by **A**, the preceding equation then generalizes to

$$\mathbf{A} = \mathbf{R} + \mathbf{QA}$$

or

$$\begin{aligned} \mathbf{A} &= (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R} \\ &= \mathbf{MR} \end{aligned}$$

The following example illustrates the computations.

Example 4.28:

A manager of one section of a plant has different employees working at level I and at level II. New employees may enter his section at either level. At the end of each year, the performance of each employee is evaluated; employees can be reassigned to their level I or II jobs, terminated, or promoted to level III, in which case they never go back to I or II. The manager can keep track of employee movement as a Markov chain. The absorbing states are termination (S_1) and employment at level III (S_2); the nonabsorbing states are employment at level 1 (S_3) and employment at level II (S_4). Records over a long period of time indicate that the following is a reasonable assignment of probabilities:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.2 & 0.1 & 0.2 & 0.5 \\ 0.1 & 0.3 & 0.1 & 0.5 \end{bmatrix}$$

Thus, if an employee enters at a level I job, the probability is 0.5 that she will jump to level II work at the end of the year, but the probability is 0.2 that she will be terminated.

1. Find the expected number of evaluations an employee must go through in this section.
2. Find the probabilities of being terminated or promoted to level III eventually.

Solution:

1. For the **P** matrix,

$$R = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} 0.2 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}$$

Thus,

$$I - Q = \begin{bmatrix} 0.8 & -0.5 \\ -0.1 & 0.5 \end{bmatrix}$$

and

$$M = (I - Q)^{-1} = \begin{bmatrix} 10/7 & 10/7 \\ 2/7 & 16/7 \end{bmatrix} = \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix}$$

It follows that

$$m_3 = \frac{20}{7} \text{ and } m_4 = \frac{18}{7}$$

In other words, a new employee in this section can expect to remain there through 20/7 evaluation periods if she enters at level I, whereas she can expect to remain there through 18/7 evaluations if she enters at level II.

2. The fact that

$$A = MR = \begin{bmatrix} 3/7 & 4/7 \\ 2/7 & 5/7 \end{bmatrix} = \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}$$

implies that an employee entering at level I has a probability of 4/7 of reaching level III, whereas an employee entering at level II has a probability of 5/7 of reaching level III. The probabilities of termination at levels I and II are therefore 3/7 and 2/7, respectively.

Example 4.29:

Continuing the genetics problem in Example 4.27, suppose that an individual of unknown genetic makeup is mated with a known homozygous dominant (GG) individual. The matrix of transition probabilities for the first-generation offspring then becomes

$$P = \begin{array}{c} \\ d \\ h \\ r \end{array} \begin{array}{c} \begin{array}{ccc} d & h & r \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

which has one absorbing state. Find the mean number of generations until all offspring become dominant.

Solution:

In the notation used earlier,

$$Q = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix}$$

$$I - Q = \begin{bmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{bmatrix}$$

and

$$(I - Q)^{-1} = \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix} = M$$

Thus, if the unknown is heterozygous, we should expect all offspring to be dominant after two generations. If the unknown is homozygous recessive, we should expect all offspring to be dominant after three generations. Notice that

$$A = MR = \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which simply indicates that we are guaranteed to reach the fully dominant state eventually, no matter what the genetic makeup of the unknown parent may be.

Exercises

4.111. A certain city prides itself on having sunny days. If it rains one day, there is a 90% chance that it will be sunny the next day. If it is sunny one day, there is a 30% chance that it will rain the following day. (Assume that there are only sunny or rainy days.) Does the city have sunny days most of the time? In the long run, what fraction of all days are sunny?

4.112. Suppose a car rental agency in a large city has three locations: downtown location (labeled A), airport location (labeled B), and a hotel location (labeled C). The agency has a group of delivery drivers to serve all three locations. Of the calls to the Downtown location, 30% are delivered in downtown area, 30% are delivered to the airport, and 40% are delivered to the hotel. Of the calls to the airport location, 40% are delivered in downtown area, 40% are delivered to the airport, and 20% are delivered to the hotel. Of the calls to the hotel location, 50% are delivered in the downtown area, 30% are delivered to the airport area, and 20% are delivered to the hotel area. After making a delivery, a driver goes to the nearest location to make the next delivery. This way, the location of a specific driver is determined only by his or her previous location.
a. Give the transition matrix.
b. Find the probability that a driver who begins in the downtown location will be at the hotel after two deliveries.
c. In the long run, what fraction of the total number of stops does a driver make at each of the three locations?

4.113. For a Markov chain, show that $\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\mathbf{P}$

4.114. Suppose that a particle moves in unit steps along a straight line. At each step, the particle either remains where it is, moves one step to the right, or moves one step to the left. The line along which the particle moves has barriers at 0 and at b, a positive integer, and the particle only moves between these barriers; it is absorbed if it lands on either barrier. Now, suppose that the particle moves to the right with probability p and to the left with probability $1 - p = q$.
a. Set up the general form of the transition matrix for a particle in this system.
b. For the case $b = 3$ and $p \neq q$, show that the absorption probabilities are as follows:

$$a_{10} = \frac{\left(\frac{q}{p}\right) - \left(\frac{q}{p}\right)^3}{1 - \left(\frac{q}{p}\right)^3}$$

$$a_{20} = \frac{\left(\frac{q}{p}\right)^2 - \left(\frac{q}{p}\right)^3}{1 - \left(\frac{q}{p}\right)^3}$$

c. In general, it can be shown that

$$a_{10} = \frac{\left(\frac{q}{p}\right)^1 - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}, \qquad p \neq q$$

By taking the limit of a_{j0} as $p \to ½$, show that

$$a_{j0} = \frac{b-j}{b}, \qquad p = q$$

d. For the case $b = 3$, find an expression for the mean time to absorption from state j, with $p \neq q$. Can you generalize this result?

4.115. Suppose that n white balls and n black balls are placed in two urns so that each urn contains n balls. A ball is randomly selected from each urn and placed in the opposite urn. (This is one possible model for the diffusion of gases.)
a. Number the urns 1 and 2. The state of the system is the number of black balls in urn 1. Show that the transition probabilities are given by the following quantities:

$$p_{jj-1} = \left(\frac{j}{n}\right)^2, \qquad j > 0$$

$$p_{jj} = \frac{2j(n-j)}{n^2}$$

$$p_{jj+1} = \left(\frac{n-j}{n}\right)^2, \qquad j < n$$

$$p_{jk} = 0, \qquad otherwise$$

b. After many transitions, show that the stationary distribution is satisfied by

$$p_j = \frac{\binom{n}{j}^2}{\binom{2n}{n}}$$

Give an intuitive argument as to why this looks like a reasonable answer.

4.116. Suppose that two friends, A and B, toss a balanced coin. If the coin comes up heads, A wins \$1 from B. If it comes up tails, B wins \$1 from A. The game ends only when one player has all the other's money. If A starts with \$1 and B with \$3, find the expected duration of the game and the probability that A will win.

4.12. Activities for Students: Simulation

Computers lend themselves nicely to use in the area of probability. Not only can computers be used to calculate probabilities, but they can also be used to simulate random ariables from specified probability distributions. A simulation, performed on the computer, permits the user to analyze both theoretical and applied problems. A simujlated model attempts to copyu the behavior of a situation under consideration; practical applications include models of inventory control problems, queuing systems, production lines, medical systems, and flight patterns of major jets. Simulation can also be used to determine the behavior of a complicated random variable whose precise probability distribution function is difficult to evaluate mathematically.

Generating observations from a probability distribution is based on random numbers on [0, 1]. A random number R_i on the interval [0, 1] is chosen, on the condition that each number between 0 and 1 has the same probability of being selected. A sequence of numbers that appears to follow a certain pattern or trend should not be considered random. Most computer languages have built-in random generators that give a number on [0, 1]. If a built-in generator is not available, algorithms are available for setting up one. One basic technique used is a congruential method, such as a linear congruential generator (LCG). The generator most commonly used is

$$x_i = (ax_{i-1} + c)(\mathrm{mod}\, m), \qquad i = 1,\ 2, \ldots$$

Where x_i, a, c, and m are integers and $0 \leq x_i < m$. If $c = 0$, this method is called a multiplicative-congruential sequence. The values of x_i generated by the LCG are contained in the interval $[0, m)$. To obtain a value in [0, 1), we evaluate x_i/m.

A random number generator should have certain characteristics, such as the following:

1. The numbers produced (x_i/m) should appear to be distributed uniformly on [0, 1]; that is, the probability function for the numbers should be constant over the interaval [0, 1]. The numbers should also be independent of each other. Any statistical tests can be applied to check for uniformity and independence, such as comparing the number of occurrences of a digit in a sequence with the expected number of occurrences for that digit.

2. The random number generator should have a long, full period; that is, the sequence of numbers should not begin repeating or "cycling" too quickly. A sequence has a full period (p) if $p = m$. It has been shown that an LCG has a full period if and only if the following are true (Kennedy and Gentle 1980, p. 137).
a. The only positive integer that (exactly) divides both m and c is 1.
b. If q is a prime number that divides m, then q divides $a-1$.
c. If 4 divides m, then 4 divides $a-1$.

This implies that, if c is odd and $a-1$ is divisible by 4, when a full-period generator is used, x_0 can be any integer between 0 and $m-1$ without affecting the generator's period. A good choice for m is 2^b, where b is the number of bits. Based on the just-mentioned relationships of a, c, and m, the following values have been found to be satisfactory for use on a microcomputer: $a = 25{,}173$; $c = 13{,}849$; $m = 2^{16} = 65{,}536$ (Yang and Robinson 1986, p. 6).

3. It is beneficial to be able to reuse the same random numbers in a different simulation run. The preceding LCG has this ability, because all values of x_i are determined by the initial value (seed) x_0. Because of this, the computed random numbers are referred to as pseudo-random numbers. Even though these numbers are not truly random, careful selection of a, c, m and x_0 will yield values for x_i that behave as random numbers and pass the appropriate statistical tests, as described in characteristic 1.

4. The generator should be efficient—that is, fast and in need of little storage.
Given that a random number R_i on [0, 1] can be generated, we will now consider a brief description of generating discrete random variables for the distributions discussed in this chapter.

4.12.1 Bernoulli Distribution

Let p represent the probability of success. If $R_i \le p$, then $X_i = 1$; otherwise, $X_i = 0$.

4.12.2 Binomial Distribution

A binomial random variable X_i can be expressed as the sum of independent Bernoulli random variables Y_j; that is $X_i = \sum Y_j$, where $j = 1, 2, \ldots, n$. Thus to simulate X_i witn parameters n and p, we simulate n Bernoulli random variables, as stated previously, X_i is equal to the sum of the n Bernoulli variables.

4.12.3 Geometric Distribution

Let X_i represent the number of failures prior to the first success, with p being probability of success, $X_i = m$, where m is the number of R_i's generated prior to generating an R_i such that $R_i \le p$.

4.12.4 Negative Binomial Distribution

Let X_i represent the number of trials prior to the rth success, with p being the probability of success. A negative binomial random variable X_i can be expressed as the sum of r independent geometric random variables Y_j; that is, $X_i = \sum Y_j$, where $j = 1, 2, \ldots, r$. Thus, to simulate X_i with parameter p, we simulate r geometric random variables, as stated previously. X_i is equal to the sum of the r geometric random variables.

4.12.5 Poisson Distribution
Generating Poisson random variables will be discussed at the end of Chapter 5.

Let us consider some simple examples of possible uses for simulating discrete random variables. Suppose that n_1 items are to be inspected from on production line and that n_2 items are to be inspected from another. Let p_1 represent the probability of a defective from line 1, and let p_2 represent the probability of a defective from line 2. Let X be a binomial random variable with parameters n_1 and p_1. Let Y be a binomial random variable with parameters n_2 and p_2. A variable of interest is W, which represents the total of defective items observed in both production lines. Let $W = X + Y$. Unless $p_1 = p_2$, the distribution of W will not be binomial. To see how the distribution of W will behave, we can perform a simulation. Useful information could be obtained from the simulation by looking at a histogram of the values of W_i generated and considering the values X and Y; X is binomial with $n_1 = 7$, $p_1 = 0.2$; and Y is binomial with $n_2 = 8$, $p_2 = 0.6$. Defining $W = X + Y$, a simulation produced the histogram shown in Simulation 1.

The sample mean was 6.2, with a sample standard deviation of 1.76. In Chapter 6, we will be able to show that these values are very close to the expected values of $\mu_W = 6.2$ and $\sigma_W = 1.74$. (This calculation will make sense after we discuss the linear function of random variables.) From the histogram, we see that the probability that the total number of defective items is at least 9 is given by 0.09.

Another example of interest might be the coupon-collector problem, which incorporates the geometric distribution. Suppose that there are n distinct colors of coupons. We assume that, each time someone obtains a coupon, it is equally likely to be any one of the n colors and the selection of the coupon is independent of any previously obtained coupon. Suppose that an individual can redeem a set of coupons for a prize if each possible color coupon is represented in the set. We define the random variable X as representing the total number of coupons that must be selected to complete a set of each color coupon at random. Questions of interest might include the following:

1. What is the expected number of coupons needed in order to obtain this complete set; that is, what is $E(X)$?
2. What is the standard deviation of X?
3. What is the probability that one must select at most x coupons to obtain this complete set?

Instead of answering these questions by deriving the distribution function of X, one might try simulation. Two simulations (Simulations 2 and 3) follow. The first histogram represents a simulation where n, the number of different color coupons, is equal to 5.

The sample mean in this case was computed to be 11.06, with a sample standard deviation of 4.65. Suppose that one is interested in finding $P(X \leq 10)$. Using the results of the simulations, the relatie frequency probability is given as 0.555. It might also be noted that, from this simulation, the largest number of coupons needed to obtain the complete set was 31.

4.12.6. Binomial Proportions

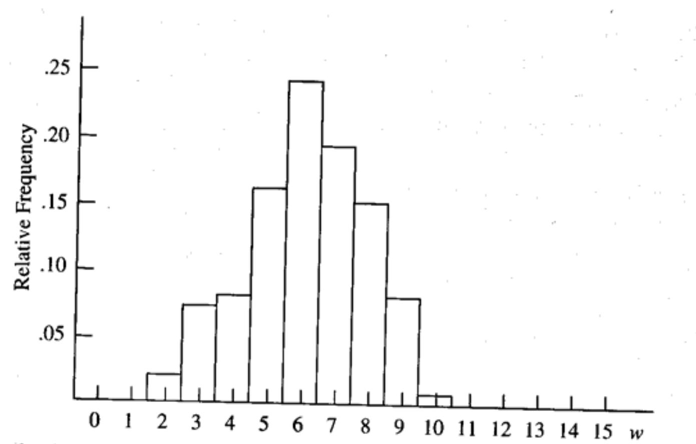
Count data are often more conveniently discussed as a proportion—as in the proportion of inspected cars that have point defects, or the proportion of sampled voters who favor a certain candidate. For a binomial random variable X on n trials, and a probability p of success on any one trial, the proportion of successes is X/n.

Generate 100 values of X/n and plot them on a line plot or histogram for each of the following cases:

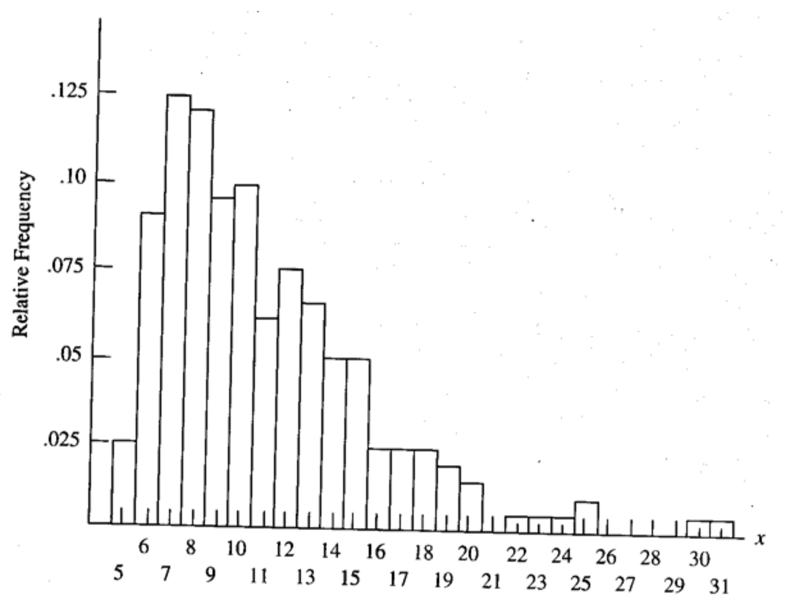
1. $n = 20,\ p = 0.2$
2. $n = 40,\ p = 0.2$
3. $n = 20,\ p = 0.5$
4. $n = 40,\ p = 0.5$

Compare the four plots, and discuss any patterns you observe in their centering, variation, and symmetry. Can you propose a formula for the variance of X/n?

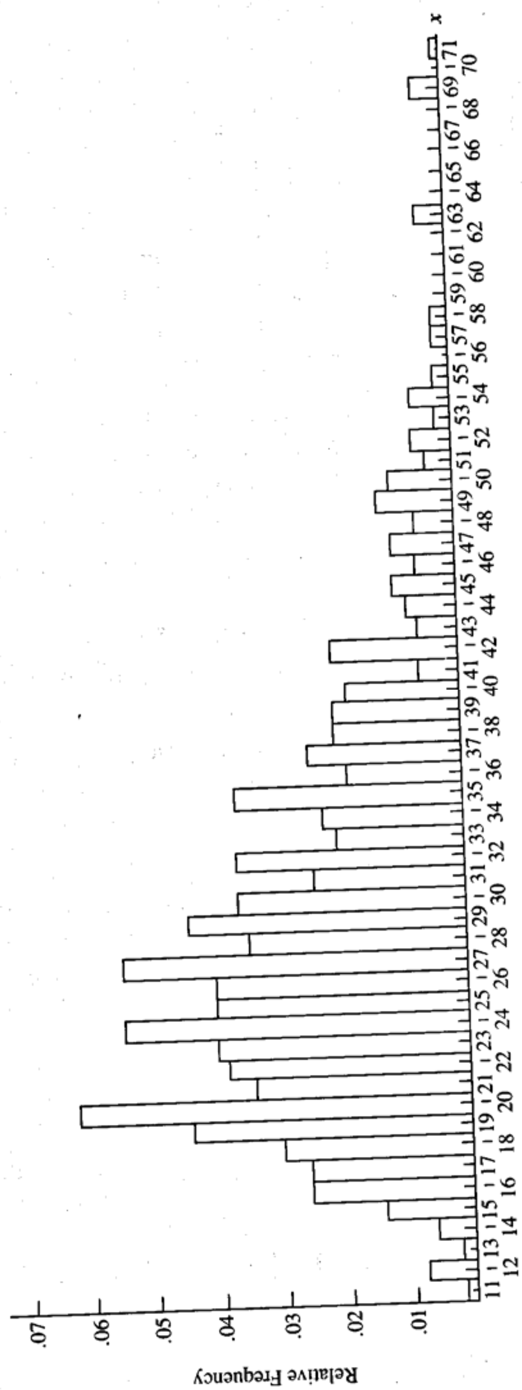
Simulation 1



Simulation 2



Simulation 3



4.12.7. Waiting for Blood

A local blood bank knows that about 40% of its donors have A^+ blood. It needs $k = 3$ A^+ donors today. Generate a distribution of values for X, the number of donors that must be tested sequentially to find the three A^+ donors needed. What are the approximate expected value and standard deviation of X from your data? Do these values agree with the theory? What is the estimated probability that the blood bank must test 10 or more people to find the three A^+ donors? How will the answers to these questions change if $k = 4$?

4.13. Summary

The outcomes of interest in most investigations involving random events are numerical. The simplest numbered outcomes to model are counts, such as the number of nonconforming parts in a shipment, the number of sunny days in a month, or the number of water samples that contain a pollutant. One amazing result of probability theory is the fact that a small number of theoretical distributions can cover a wide array of applications. Six of the most useful discrete probability distributions are introduced in this chapter.

The *Bernoulli* random variable is simply an indicator random variable; it uses a numerical code to indicate the presence or absence of a characteristic.

The *binomial* random variable counts the number of "successes" among a fixed number n of independent events, each with the same probability p of success.

The *geometric* random variable counts the number of failures one obtains prior needs to conduct sequentially until the first "success" is seen when conducting independent Bernoulli trials, each with the probability p of success.

The *negative binomial* random variable counts the number of failures observed when Bernoulli trials, each with the probability p of success, are conducted sequentiallyt until the rth success is obtained. The negative binomial may be derived from several other models and is often considered as a model for discrete count data if the variance exceeds the mean.

The *Poisson* random variable arises from counts in a restricted domain of time, area, or volume, and is most useful for counting fairly rare outcomes.

The *hypergeometric* random variable counts the number of "successes in sampling from a finite population, which makes the sequential selections dependent on one another.

These theoretical distributions serve as models for read daa that might arise in our quest to improve a process. Each involves assumptions that should be checked carefully before the distribution is applied.

Supplementary Exercises

4.117. Construct probability histograms for the binomial probability distribution for $n = 7$, and $p = 0.2$, 0.5, and 0.8. Notice the symmetry for $p = 0.5$ and the direction of skewness for $p = 0.2$ and $p = 0.8$.

4.118. Construct a probability histogram for the binomial probability distribution for $n = 20$ and $p = 0.5$. Notice that almost all of the probability falls in the interval $5 \leq y \leq 15$.

4.119. The probability that a single field radar set will detect an enemy plane is 0.8. Assume that the sets operate independently of each other.
a. If we have five radar sets, what is the probability that exactly four sets will detect the plane?
b. At least one set will detect the plane?

4.120. Suppose that the four engines of a commercial aircraft were arranged to operate independently and that the probability of in-flight failure of a single engine is 0.01. What are the probabilities that, on a given flight, the following events occur?
a. No failures are observed
b. No more than one failure is observed
c. All engines fail

4.121. Sampling for defectives from among large lots of a manufactured product yields a number of defectives Y that follows a binomial probability distribution. A sampling plan consists of specifying the number n of items to be included in a sample and an acceptance number a. The lot is accepted if $Y \leq a$, and is rejected if $Y > a$. Let p denote the proportion of defectives in the lot. For $n = 5$ and $a = 0$, calculate the probability of lot acceptance if the following lot proportions of defectives exist.
a. $p = 0$
b. $p = 0.3$
c. $p = 1.0$
d. $p = 0.1$
e. $p = 0.5$
A graph showing the probability of lot acceptance as a function of lot fraction defective is called the operating characteristic curve for the sample plan. Construct this curve for the plan $n = 5$, $a = 0$. Notice that a sampling plan is an example of statistical inference. Accepting or rejecting a lot based on information contained in the sample is equivalent to concluding that the lot is either good or bad, respectively. "Good" implies that a low fraction of items are defective and, therefore, that the lot is suitable for shipment.

4.122. Refer to exercise 4.121. Construct the operating characteristic curve for a sampling plan with the following values
a. $n = 10$, $a = 0$
b. $n = 10$, $a = 1$
c. $n = 10$, $a = 2$
For each, calculate P(lot acceptance) for $p = 0.05$, 0.1, 0.3, 0.5, and 1.0. Our intuition suggests that sampling plan (a) would be much less likely to accept bad lots than would plan (b) or plan (c). A visual comparison of the operating characteristic curves will confirm this supposition.

4.123. Refer to exercise 4.121. A quality control engineer wishes to study two alternative sample plans: $n = 5$, $a = 1$; and $n = 25$, $a = 5$. On a sheet of graph paper, construct the operating characteristic curves for both plans; make use of acceptance probabilities at $p = 0.05$, 0.10, 0.20, 0.30, and 0.40 in each case.

a. If you were a seller producing lots whose fraction of defective items ranged from $p = 0$ to $p = 0.10$, which of the two sampling plans would you prefer?
b. If you were a buyer wishing to be protected against accepting lots with a fraction defective exceed $p = 0.30$, which of the two sampling plans would you prefer?

4.124. At an archaeological site that was an ancient swamp, the bones from 22 Tyrannosaurus rex skeletons have been unearthed. The bones do not show any sign of disease or malformation. It is thought that these animals wandered into a deep area of the swamp and became trapped in the swamp bottom. The 22 right femur bones (thigh bones) were located and 5 of these right femurs are to be randomly selected without replacement for DNA testing to determine gender. Let X be the number out of the 5 selected right femurs that are from males.
a. Based on how these bones were sampled, explain why the probability distribution of X is *not* binomial.
b. Would the binomial distribution provide a good approximation to the probability distribution of X? Justify your answer.
c. Suppose that the group of 22 Tyrannasaurus whose remains were found in the swamp had been made up of 11 males and 11 females. What is the probability that all 5 in the sample to be tested are male? (Do not use the binomial approximation.)
d. The DNA testing revealed that all 5 femurs tested were from males. Based on this result and your answer from part (b), do you think that males and females were equally represented in the group of 22 brontosaurs stuck in the swamp? Explain.

4.125. For a certain section of a pine forest, the number Y of diseased trees per acre has a Poisson distribution with mean $\lambda = 10$. To treat the trees, spraying equipment must be rented for \$50. The diseased trees are sprayed with an insecticide at a cost of \$5 per tree. Let C denote the total spraying cost for a randomly selected acre.
a. Find the expected value and the standard deviation for C.
b. Within what interval would you expect C to lie with a probability of at least 0.80?

4.126. In checking river water samples for bacteria, a researcher places water in a culture medium to grow colonies of certain bacteria, if present. The number of colonies per dish averages fifteen for water samples from a certain river.
a. Find the probability that the next dish observed will have at least ten colonies.
b. Find the mean and the standard deviation of the number of colonies per dish.
c. Without calculating exact Poisson probabilities, find an interval in which at least 75%of the colony count measurements should lie.

4.127. The number of vehicles passing a specified point on a highway averages eight per minute.
a. Find the probability that at least 15 vehicles will pass this point in the next minute.
b. Find the probability that at least 15 vehicles will pass this point in the next two minutes.
c. What assumptions must you make for you answers in (a) and (b) to be valid?

4.128. A production line often produces a variable number N of items each day. Suppose that each item produced has the same probability p of not conforming to manufacturing standards. If N has a Poisson distribution with mean λ, then the number of nonconforming items in one day's production Y has a Poisson distribution with mean λp. The average number of resistors produced by a facility in one day has a Poisson distribution, with a mean of 100. Typically, 5% of the resistors produced do not meet specifications.
a. Find the expected number of resistors that will not meet specifications on a given day.
b. Find the probability that all resistors will meet the specifications on a given day.
c. Find the probability that more than five resistors will fail to meet specifications on a given day.

4.129. The mean number of customers arriving in a bank during a randomly selected hour is 4. The bank manager is considering reducing the number of tellers, but she wants to be sure that lines do not get too long. She decides that, if no more than two customers come in during a 15-minute period, two tellers (instead of the current 3) will be sufficient.
a. What is the probability no more than 2 customers will come in the bank during a randomly selected 15-minute period?
b. What is the probability that more than two customers will come in during two consecutive 15-minute periods?
c. The manager records the number of customers coming in during 15-minute time periods until she observes a time period during which more than two customers arrive. Eight time periods have been recorded, each with two or fewer customers arriving in each. What is the probability that more than 14 time periods will be observed before having more than two customers arriving during a timer 15-minute period.

4.130. An interviewer is given a list of potential people she can interview. Suppose that the interviewer needs to interview 5 people and that each person independently agrees to be interviewed with probability 0.6. Let X be the number of people who refuse the interview before she obtains her necessary number of interviews.
a. What is the probability that no more than two people will refuse the interview before she finds 5 people to interview?
b. What is the expected value and variance of the number of people who refuse the interview before obtaining 5 people to interview?

4.131. Three men flip coins to see who pays for coffee. If all three match (all heads or all tails), they flip again. Otherwise, the "odd man" pays for coffee.
a. What is the probability that they will need to flip the coins more than once?
b. What is the probability they will need to toss the coins more than three times?
c. Suppose the men have flipped the coins three times and all three times they matched. What is the probability that they will need to flip the coins more than three more times (more than 6 times total)?

4.132. A certain type of bacteria cell divides at a constant rate λ over time. Thus, the probability that a particular cell will divide in a small interval of time t is approximately λt. Given that a population starts out at time zero with k cells of this type, and cell divisions are independent of one another, the size of the population at time t, $Y(t)$, has the probability distribution

$$P[Y(t)=n]=\binom{n-1}{k-1}e^{-\lambda kt}\left(1-e^{-\lambda t}\right)^{n-k}, \qquad n=k,\ k+1,\ldots$$

a. Find the expected value of $Y(t)$ in terms of λ and t.
b. If, for a certain type of bacteria cell, $\lambda = 0.1$ per second, and the population starts out with two cells at time zero, find the expected population size after 5 seconds.

4.133. In a certain region, the probability of at least one child contracting malaria during a given week is 0.2. Find the average number of cases during a four-week period, assuming that a person contracting malaria is independent of another person contracting malaria.

4.134. The probability that any one vehicle will turn left at a particular intersection is 0.2. The left-turn lane at this intersection has room for three vehicles. If five vehicles arrive at this intersection while the light is red, find the probability that the left-turn lane will hold all of the vehicles that want to turn left.

4.135. Referring to Exercise 4.134, find the probability that six cars must arrive at the intersection while the light is red to fill up the left-turn lane.

4.136. For any probability $p(y)$, $\sum_y p(y)=1$ if the sum is taken over all possible values y that the random variable in question can assume. Show that this is true for the following distributions.
a. The binomial distribution
b. The geometric distribution
c. The Poisson distribution

4.137. During World War I, the British government established the Industrial Fatigue Research Board (IFRB), later known as the Industrial Health Research Board (IHRB) (Haight 2001). The board was created because of concern of the large number of accidental deaths and injuries in the British war production industries. One of the data sets they considered was the number of accidents experienced by women working on 6-inch shells during the period February 13, 1918, to March 20, 1918. These are displayed in the table below. Thus, 447 women had no accidents during this time period, but two had at least five accidents.

Number of Accidents	Frequency Observed
0	447
1	132
2	42
3	21
4	3
5 or more	2

a. Find the average number of accidents a woman had during this time period. (Assume all observations in the category "5 or more" are exactly 5.)
b. After the work of von Bortkiewicz (see Section 4.8.3), the Poisson had been applied to a large number of random phenomena and, with few exceptions, had been found to describe the data well. This had led the Poisson distribution to be called the "random distribution," a term that is still found in the literature. Thus, the mathematicians at the IFRB began by modeling these data using the Poisson distribution. Find the expected number of women having 0, 1, 2, 3, 4, and ≥ 5 accidents using the mean found in (a) and the Poisson distribution. How well do you think this model describes the data?
c. Greenwood and Woods (1919) suggested fitting a negative binomial distribution to these data. Find the expected number of women having 0, 1, 2, 3, 4, and ≥ 5 accidents using the mean found in (a) and the geometric (negative binomial with $r = 1$) distribution. How well do you think this model describes the data?

Historical note: Researchers were puzzled as to why the negative binomial fit better than the Poisson until, in 1920, Greenwood and Yule suggested the following model. Suppose that the probability any given woman will have an accident is distributed according to a Poisson distribution with mean λ. However, λ varies from woman to woman according to a gamma distribution (see Chapter 5). Then the number of accidents would have a negative binomial distribution. The value λ of lambda associated with a woman was called her "accident proneness."

4.138. The Department of Transportation's Federal Auto Insurance and Compensation Study was based on a random sample of 7,842 California licensed drivers (Ferreira, 1971). The number of accidents in which each was involved from November, 1959, to February, 1968, was determined. The summary results are given in the table below.

Number of Accidents	Frequency Observed
0	5,147
1	1,849
2	595
3	167
4	54
5	14
6 or more	6

a. Find the average number of accidents a California licensed driver had during this time period. (Assume all observations in the category "6 or more" are exactly 6.)
b. Find the expected number of drivers being involved in 0, 1, 2, 3, 4, 5 and ≥ 6 accidents using the mean found in (a) and the Poisson distribution. How well do you think this model describes the data?
c. Find the expected number of women having 0, 1, 2, 3, 4, 5 and ≥ 6 accidents using the mean found in (a) and the geometric (negative binomial with $r = 1$) distribution. How well do you think this model describes the data?

Note: In both this and Exercise 4.137, a better fit of the negative binomial can be found by using a non-integer value for r. Although we continue to restrict our attention to only integer values, the negative binomial distribution is well defined for any real value $r > 0$.

4.139. The supply office for a large construction firm has three welding units of Brand A in stock. If a welding unit is requested, the probability is 0.7 that the request will be for this particular brand. On a typical day, five requests for welding units come to the office. Find the probability that all three Brand A units will be in use on that day.

4.140. Refer to Exercise 4.139. If the supply office also stocks three welding units that are not Brand A, find the probability that exactly one of these units will be left immediately after the third Brand A unit is requested.

4.141. In the game Lotto 6-49, six numbers are randomly chosen, without replacement, from 1 to 49. A player who matches all six numbers, in any order, wins the jackpot.
a. What is the probability of winning any given jackpot with one game ticket?
b. If a game ticket costs \$1.00, what is the expected winnings from playing Lotto 6-49 once.
c. Suppose a person buys one Lotto 6-49 ticket each week for a hundred years. Assuming all years have 52 weeks, what is the probability of winning at least one jackpot during this time? (Hint: Use a Poisson approximation)
d. Given the setting in (c), what is the expected winnings over 100 years?

4.142. The probability of a customer's arriving at a grocery service counter in any one second equals 0.1. Assume that customers arrive in a random stream and, hence, that the arrival at any one second is independent of any other arrival.
a. Find the probability that the first arrival will occur during the third 1-second interval.
b. Find the probability that the first arrival will not occur until at least the third 1-second interval.
c. Find the probability that no arrivals will occur in the first 5 seconds
d. Find the probability that at least three people will arrive in the first 5 seconds.

4.143. Of a population of consumers, 60% are reputed to prefer a particular brand, A, of toothpaste. If a group of consumers are interviewed, find the probability of the following events.
a. exactly five people are interviewed before encountering a consumer who prefers brand A
b. at least five people are interviewed before encountering a consumer who prefers brand A

4.144. The mean number of automobiles entering a mountain tunnel per 2-minute period is one. If an excessive number of cars enter the tunnel during a brief period of time, the result is a hazardous situation.
a. Find the probability that the number of autos entering the tunnel during a 2-minute period exceeds three.

b. Assume that the tunnel is observed during ten 2-minute intervals, thus giving ten independent observations, $Y_1, Y_2, \ldots Y_{10}$, on a Poisson random variable. Find the probability that $Y > 3$ during at least one of the ten 2-minute intervals.

4.145. Suppose that 10% of a brand of Mp3 players will fail before their guarantee has expired. Suppose 1000 players are sold this month, and let Y denote the number that will not fail during the guarantee period.
a. Find the expected value and variance of Y.
b. Within what limit would Y be expected to fail? (Hint: Use Tchebysheff's Theorerm.)

4.146. a. Consider a binomial experiment for $n = 20$ and $p = 0.05$. Calculate the binomial probabilities for $Y = 0, 1, 2, 3$, and 4.
b. Calculate the same probabilities, but this time use the Poisson approximation with $\lambda = np$. Compare the two results.

4.147. The manufacturer of a low-calorie dairy drink wishes to compare the taste appeal of a new formula (B) with that of the standard formula (A). Each of four judges is given three glasses in random order; two containing formula A and the other containing formula B. Each judge is asked to choose which glass he most enjoyed. Suppose that the two formulas are equally attractive. Let Y be the number of judges stating a preference for the new formula.
a. Find the probability function for Y.
b. What is the probability that at least three of the four judges will state a preference for the new formula?
c. Find the expected value of Y.
d. Find the variance of Y.

4.148. Show that the hypergeometric probability function approaches the binomial in the limit as $N \to \infty$ and as $p = r/N$ remains constant; that is, show that

$$\lim_{\substack{N\to\infty \\ r\to\infty}} = \frac{\binom{r}{y}\binom{N-r}{n-r}}{\binom{N}{n}} = \binom{n}{y} p^y q^{n-y}$$

for constant $p = r/N$.

4.149. A lot of $N = 100$ industrial products contains 40 defectives. Let Y be the number of defectives in a random sample of size 20. Find $p(10)$ using the following distributions.
a. The hypergeometric probability distribution
b. The binomial probability distribution
Is N large enough so that the binomial probability function provides a good approximation to the hypergeometric probability function?

4.150. For simplicity, let us assume that there are two kinds of drivers. The safe drivers, who constitute 70% of the population, have a probability of 0.1 of causing an accident in a year. The rest of the population are accident makers, who have a probability of 0.5 of causing an accident in a year. The insurance premium is \$1800 times one's probability of causing an accident in the following year. A new subscriber has caused an accident during the first year. What should her insurance premium be for the next year?

4.151. A merchant stocks a certain perishable item. He knows that on any given day he will have a demand for two, three, or four of these items, with probabilities 0.2, 0.3, and 0.5, respectively. He buys the items for \$1.00 each and sells them for \$1.20 each. Any items left at the end of the day represent a total loss. How many items should the merchant stock to maximize his expected daily profit?

4.152. It is known that 5% of the population has disease A, which can be discovered by means of a blood test. Suppose that N (a large number) people are tested. This can be done in two ways:
1. Each person is tested separately.
2. The blood sample of k people are pooled together and analyzed. (Assume that $N = nk$, with n an integer.) If the test is negative, all of the persons in the pool are healthy (that is, just this one test is needed). If the test is positive, each of the k persons must be tested separately (that is, a total of $k + 1$ tests are needed.)
a. For fixed k, what is the expected number of tests needed in method (2)?
b. Find the value for k that will minimize the expected number of tests in method (2).
c. How many tests does part (b) save in comparison with part (a)?

4.153. Four possible winning numbers for a lottery—AC-6732, FK-1972, OJ-8201, and PM-1182—are given to you. You will win a prize if one of your numbers marches one of the winning numbers. You are told that there is one first prize of \$250,000; two second prizes of \$75,000, and then third prizes of \$1000 each. All you have to do is mail the coupon back; no purchase is required. From the structure of the numbers you have received, it is obvious that the entire list consists of all the permutations of two letters from the alphabet, followed by four digits. Is the coupon worth mailing back for 45 cents postage?

4.154. For a discrete random variable X taking on values 0, 1, 2, …, show that
$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

References

Bortkewitsch. 1898. *Das Gasetz der Kleinen Zahlen*. Leipzig.

Boswell, M.T. and G.P. Patil. 1970. Chance mechanisms generating the negative binomial distribution. In G.P. Patil (*ed.*). *Random Counts in Scientific Work*. Vol 1. University Park: Pennsylvania State University.

Chu, Singfat. 2003. Using soccer goals to motivate the Poisson process. *INFORMS Transactions on Education* 3(2): http://ite.pubs.informs.org/Vol3No2/Chu/.

Eves, Howard. (1969). *Introduction to the History of Mathematics*, 3rd Edition. New York: Holt, Rinehart and Winston.

David, F. N. (1955). Studies in the history of probability and statistics. *Biometrika* 52: 1-15.

Ferreira, Joseph Jr. 1976. *Designing Equitable Merit Rating Plans*. Working Paper #OR-057-76. Operations Research Center, Massachusetts Institute of Technology.

Folks, J. L. (1981). *Ideas of Statistics*. New York: John Wiley & Sons.

Goranson, U.G. and J. Hall. (1980). Airworthiness of long life jet transport structures. *Aeronautical Journal* 84: 374-385.

Greenwood, M. and H.M. Woods. 1919. *The Incidence of Industrial Accidents upon Individuals with Specific Reference to Multiple Accidents*. Industrial Fatigue Research Board. Report No. 4. London.

Greenwood, M. and C.V. Yule. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society* 89: 255-279.

Haight, Frank A. 2001. *Accident Proneness: The History of an Idea*. Technical Report #UCIITS-WP-01-4. Institute of Transportation Studies, University of California, Irvine.

Sakamoto, C.M. 1973. Application of the Poisson and Negative Binomial models to thunderstorms and hail days probabilities in Nevada. *Monthly Weather Review* 101: 350-355.

Schmuland, B. 2001. Shark attacks and the Poisson approximation, *π in the Sky*, Issue 4, http://www.pims.math.ca/pi/issue4/page12-14.pdf.

Kemeny, J.G. and J.L. Snell. 1983. *Finite Markov Chains*. Springer-Verlag.

Kennedy, William and James Gentle. 1980. *Statistical Computing*. New York: Marcel Dekker, Inc.

Stigler, Stephen M. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Student. 1907. On the error of counting with a haemacytometer. *Biometrika* 5: 351-360.

Whitaker, Lucy. 1915. On the Poisson law of large numbers. *Biometrika* 10: 36-71.

Yang, Mark C.K. and David H. Robinson. 1986. *Understanding and Learning Statistics by Computer*. World Scientific.