Mathematical Expectation

1 Introduction

Originally, the concept of a **mathematical expectation** arose in connection with games of chance, and in its simplest form it is the product of the amount a player stands to win and the probability that he or she will win. For instance, if we hold one of 10,000 tickets in a raffle for which the grand prize is a trip worth \$4,800, our mathematical expectation is $4{,}800 \cdot \frac{1}{10{,}000} = \$0.48$. This amount will have to be interpreted in the sense of an average—altogether the 10,000 tickets pay \$4,800, or on the average $\frac{\$4{,}800}{10{,}000} = \$0.48$ per ticket.

If there is also a second prize worth \$1,200 and a third prize worth \$400, we can argue that altogether the 10,000 tickets pay $\$4{,}800 + \$1{,}200 + \$400 = \$6{,}400$, or on the average $\frac{\$6{,}400}{10{,}000} = \$0.64$ per ticket. Looking at this in a different way, we could argue that if the raffle is repeated many times, we would lose 99.97 percent of the time (or with probability 0.9997) and win each of the prizes 0.01 percent of the time (or with probability 0.0001). On the average we would thus win

$$0(0.9997) + 4{,}800(0.0001) + 1{,}200(0.0001) + 400(0.0001) = \$0.64$$

which is the sum of the products obtained by multiplying each amount by the corresponding probability.

2 The Expected Value of a Random Variable

In the illustration of the preceding section, the amount we won was a random variable, and the mathematical expectation of this random variable was the sum of the products obtained by multiplying each value of the random variable by the corresponding probability. Referring to the mathematical expectation of a random variable simply as its **expected value**, and extending the definition to the continuous case by replacing the operation of summation by integration, we thus have the following definition.

From Chapter 4 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

DEFINITION 1. EXPECTED VALUE. *If* X *is a discrete random variable and* f(x) *is the value of its probability distribution at* x*, the **expected value of X** is*

$$E(X) = \sum_x x \cdot f(x)$$

Correspondingly, if X *is a continuous random variable and* f(x) *is the value of its probability density at* x*, the **expected value of X** is*

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

In this definition it is assumed, of course, that the sum or the integral exists; otherwise, the mathematical expectation is undefined.

EXAMPLE 1

A lot of 12 television sets includes 2 with white cords. If 3 of the sets are chosen at random for shipment to a hotel, how many sets with white cords can the shipper expect to send to the hotel?

Solution

Since x of the 2 sets with white cords and $3-x$ of the 10 other sets can be chosen in $\binom{2}{x}\binom{10}{3-x}$ ways, 3 of the 12 sets can be chosen in $\binom{12}{3}$ ways, and these $\binom{12}{3}$ possibilities are presumably equiprobable, we find that the probability distribution of X, the number of sets with white cords shipped to the hotel, is given by

$$f(x) = \frac{\binom{2}{x}\binom{10}{3-x}}{\binom{12}{3}} \quad \text{for } x = 0, 1, 2$$

or, in tabular form,

x	0	1	2
$f(x)$	$\frac{6}{11}$	$\frac{9}{22}$	$\frac{1}{22}$

Now,

$$E(X) = 0 \cdot \frac{6}{11} + 1 \cdot \frac{9}{22} + 2 \cdot \frac{1}{22} = \frac{1}{2}$$

and since half a set cannot possibly be shipped, it should be clear that the term "expect" is not used in its colloquial sense. Indeed, it should be interpreted as an average pertaining to repeated shipments made under the given conditions.

EXAMPLE 2

Certain coded measurements of the pitch diameter of threads of a fitting have the probability density

$$f(x) = \begin{cases} \dfrac{4}{\pi(1+x^2)} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the expected value of this random variable.

Solution

Using Definition 1, we have

$$\begin{aligned} E(X) &= \int_0^1 x \cdot \frac{4}{\pi(1+x^2)}\,dx \\ &= \frac{4}{\pi}\int_0^1 \frac{x}{1+x^2}\,dx \\ &= \frac{\ln 4}{\pi} = 0.4413 \end{aligned}$$

There are many problems in which we are interested not only in the expected value of a random variable X, but also in the expected values of random variables related to X. Thus, we might be interested in the random variable Y, whose values are related to those of X by means of the equation $y = g(x)$; to simplify our notation, we denote this random variable by $g(X)$. For instance, $g(X)$ might be X^3 so that when X takes on the value 2, $g(X)$ takes on the value $2^3 = 8$. If we want to find the expected value of such a random variable $g(X)$, we could first determine its probability distribution or density and then use Definition 1, but generally it is easier and more straightforward to use the following theorem.

THEOREM 1. If X is a discrete random variable and $f(x)$ is the value of its probability distribution at x, the expected value of $g(X)$ is given by

$$E[g(X)] = \sum_x g(x) \cdot f(x)$$

Correspondingly, if X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of $g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx$$

Proof Since a more general proof is beyond the scope of this chapter, we shall prove this theorem here only for the case where X is discrete and has a finite range. Since $y = g(x)$ does not necessarily define a one-to-one correspondence, suppose that $g(x)$ takes on the value g_i when x takes on the values $x_{i1}, x_{i2}, \ldots, x_{in_i}$. Then, the probability that $g(X)$ will take on the value g_i is

$$P[g(X) = g_i] = \sum_{j=1}^{n_i} f(x_{ij})$$

and if $g(x)$ takes on the values $g_1, g_2, \ldots, g_m$, it follows that

$$\begin{aligned}
E[g(X)] &= \sum_{i=1}^{m} g_i \cdot P[g(X) = g_i] \\
&= \sum_{i=1}^{m} g_i \cdot \sum_{j=1}^{n_i} f(x_{ij}) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n_i} g_i \cdot f(x_{ij}) \\
&= \sum_{x} g(x) \cdot f(x)
\end{aligned}$$

where the summation extends over all values of X.

EXAMPLE 3

If X is the number of points rolled with a balanced die, find the expected value of $g(X) = 2X^2 + 1$.

Solution

Since each possible outcome has the probability $\frac{1}{6}$, we get

$$\begin{aligned}
E[g(X)] &= \sum_{x=1}^{6} (2x^2 + 1) \cdot \frac{1}{6} \\
&= (2 \cdot 1^2 + 1) \cdot \frac{1}{6} + \cdots + (2 \cdot 6^2 + 1) \cdot \frac{1}{6} \\
&= \frac{94}{3}
\end{aligned}$$

EXAMPLE 4

If X has the probability density

$$f(x) = \begin{cases} e^x & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the expected value of $g(X) = e^{3X/4}$.

Solution
According to Theorem 1, we have

$$\begin{aligned} E[e^{3X/4}] &= \int_0^\infty e^{3x/4} \cdot e^{-x}\,dx \\ &= \int_0^\infty e^{-x/4}\,dx \\ &= 4 \end{aligned}$$

The determination of mathematical expectations can often be simplified by using the following theorems, which enable us to calculate expected values from other known or easily computed expectations. Since the steps are essentially the same, some proofs will be given for either the discrete case or the continuous case; others are left for the reader as exercises.

THEOREM 2. If a and b are constants, then

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

Proof Using Theorem 1 with $g(X) = aX + b$, we get

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^{\infty} (ax+b) \cdot f(x)\,dx \\ &= a\int_{-\infty}^{\infty} x \cdot f(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx \\ &= aE(X) + b \end{aligned}$$

If we set $b = 0$ or $a = 0$, we can state the following corollaries to Theorem 2.

COROLLARY 1. If a is a constant, then

$$E(aX) = aE(X)$$

COROLLARY 2. If b is a constant, then

$$E(b) = b$$

Observe that if we write $E(b)$, the constant b may be looked upon as a random variable that always takes on the value b.

THEOREM 3. If $c_1, c_2, \ldots,$ and c_n are constants, then

$$E\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i E[g_i(X)]$$

Proof According to Theorem 1 with $g(X) = \sum_{i=1}^{n} c_i g_i(X)$, we get

$$\begin{aligned}
E\left[\sum_{i=1}^{n} c_i g_i(X)\right] &= \sum_{x}\left[\sum_{i=1}^{n} c_i g_i(x)\right] f(x) \\
&= \sum_{i=1}^{n}\sum_{x} c_i g_i(x) f(x) \\
&= \sum_{i=1}^{n} c_i \sum_{x} g_i(x) f(x) \\
&= \sum_{i=1}^{n} c_i E[g_i(X)]
\end{aligned}$$

EXAMPLE 5

Making use of the fact that

$$E(X^2) = (1^2+2^2+3^2+4^2+5^2+6^2)\cdot\frac{1}{6} = \frac{91}{6}$$

for the random variable of Example 3, rework that example.

Solution

$$E(2X^2+1) = 2E(X^2)+1 = 2\cdot\frac{91}{6}+1 = \frac{94}{3}$$

EXAMPLE 6

If the probability density of X is given by

$$f(x) = \begin{cases} 2(1-x) & \text{for } 0<x<1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) show that

$$E(X^r) = \frac{2}{(r+1)(r+2)}$$

(b) and use this result to evaluate

$$E[(2X+1)^2]$$

Solution

(a)

$$\begin{aligned}
E(X^r) &= \int_0^1 x^r\cdot 2(1-x)\,dx = 2\int_0^1 (x^r - x^{r+1})\,dx \\
&= 2\left(\frac{1}{r+1}-\frac{1}{r+2}\right) = \frac{2}{(r+1)(r+2)}
\end{aligned}$$

(b) Since $E[(2X+1)^2] = 4E(X^2)+4E(X)+1$ and substitution of $r = 1$ and $r = 2$ into the preceding formula yields $E(X) = \frac{2}{2\cdot 3} = \frac{1}{3}$ and $E(X^2) = \frac{2}{3\cdot 4} = \frac{1}{6}$, we get

$$E[(2X+1)^2] = 4\cdot\frac{1}{6}+4\cdot\frac{1}{3}+1 = 3$$

EXAMPLE 7

Show that

$$E[(aX+b)^n] = \sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^iE(X^{n-i})$$

Solution

Since $(ax+b)^n = \sum_{i=0}^{n}\binom{n}{i}(ax)^{n-i}b^i$, it follows that

$$\begin{aligned}E[(aX+b)^n] &= E\left[\sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^iX^{n-i}\right]\\ &= \sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^iE(X^{n-i})\end{aligned}$$

The concept of a mathematical expectation can easily be extended to situations involving more than one random variable. For instance, if Z is the random variable whose values are related to those of the two random variables X and Y by means of the equation $z = g(x, y)$, we can state the following theorem.

THEOREM 4. If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the expected value of $g(X, Y)$ is

$$E[g(X, Y)] = \sum_x\sum_y g(x, y)\cdot f(x, y)$$

Correspondingly, if X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y), the expected value of $g(X, Y)$ is

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)\,dx\,dy$$

Generalization of this theorem to functions of any finite number of random variables is straightforward.

EXAMPLE 8

Find the expected value of $g(X, Y) = X + Y$.

Solution

$$
\begin{aligned}
E(X+Y) &= \sum_{x=0}^{2}\sum_{y=0}^{2}(x+y)\cdot f(x,y) \\
&= (0+0)\cdot\frac{1}{6} + (0+1)\cdot\frac{2}{9} + (0+2)\cdot\frac{1}{36} + (1+0)\cdot\frac{1}{3} \\
&\quad + (1+1)\cdot\frac{1}{6} + (2+0)\cdot\frac{1}{12} \\
&= \frac{10}{9}
\end{aligned}
$$

EXAMPLE 9

If the joint probability density of X and Y is given by

$$
f(x,y) = \begin{cases} \frac{2}{7}(x+2y) & \text{for } 0 < x < 1, 1 < y < 2 \\ 0 & \text{elsewhere} \end{cases}
$$

find the expected value of $g(X, Y) = X/Y^3$.

Solution

$$
\begin{aligned}
E(X/Y^3) &= \int_1^2\int_0^1 \frac{2x(x+2y)}{7y^3}\,dx\,dy \\
&= \frac{2}{7}\int_1^2\left(\frac{1}{3y^3}+\frac{1}{y^2}\right)dy \\
&= \frac{15}{84}
\end{aligned}
$$

The following is another theorem that finds useful applications in subsequent work. It is a generalization of Theorem 3, and its proof parallels the proof of that theorem.

THEOREM 5. If $c_1, c_2, \ldots,$ and c_n are constants, then

$$
E\left[\sum_{i=1}^{n} c_i g_i(X_1, X_2, \ldots, X_k)\right] = \sum_{i=1}^{n} c_i E[g_i(X_1, X_2, \ldots, X_k)]
$$

Exercises

1. To illustrate the proof of Theorem 1, consider the random variable X, which takes on the values -2, -1, 0, 1, 2, and 3 with probabilities $f(-2)$, $f(-1)$, $f(0)$, $f(1)$, $f(2)$, and $f(3)$. If $g(X) = X^2$, find
(a) g_1, g_2, g_3, and g_4, the four possible values of $g(x)$;
(b) the probabilities $P[g(X) = g_i]$ for $i = 1, 2, 3, 4$;
(c) $E[g(X)] = \sum_{i=1}^{4} g_i \cdot P[g(X) = g_i]$, and show that it equals

$$\sum_x g(x) \cdot f(x)$$

2. Prove Theorem 2 for discrete random variables.

3. Prove Theorem 3 for continuous random variables.

4. Prove Theorem 5 for discrete random variables.

5. Given two continuous random variables X and Y, use Theorem 4 to express $E(X)$ in terms of
(a) the joint density of X and Y;
(b) the marginal density of X.

6. Find the expected value of the discrete random variable X having the probability distribution

$$f(x) = \frac{|x-2|}{7} \quad \text{for } x = -1, 0, 1, 3$$

7. Find the expected value of the random variable Y whose probability density is given by

$$f(y) = \begin{cases} \frac{1}{8}(y+1) & \text{for } 2 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

8. Find the expected value of the random variable X whose probability density is given by

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2-x & \text{for } 1 \leqq x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

9. (a) If X takes on the values 0, 1, 2, and 3 with probabilities $\frac{1}{125}$, $\frac{12}{125}$, $\frac{48}{125}$, and $\frac{64}{125}$, find $E(X)$ and $E(X^2)$.
(b) Use the results of part (a) to determine the value of $E[(3X+2)^2]$.

10. (a) If the probability density of X is given by

$$f(x) = \begin{cases} \frac{1}{x(\ln 3)} & \text{for } 1 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

find $E(X)$, $E(X^2)$, and $E(X^3)$.
(b) Use the results of part (a) to determine $E(X^3 + 2X^2 - 3X + 1)$.

11. If the probability density of X is given by

$$f(x) = \begin{cases} \frac{x}{2} & \text{for } 0 < x \leqq 1 \\ \frac{1}{2} & \text{for } 1 < x \leqq 2 \\ \frac{3-x}{2} & \text{for } 2 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

find the expected value of $g(X) = X^2 - 5X + 3$.

12. This has been intentionally omitted for this edition.

13. This has been intentionally omitted for this edition.

14. This has been intentionally omitted for this edition.

15. This has been intentionally omitted for this edition.

16. If the probability distribution of X is given by

$$f(x) = \left(\frac{1}{2}\right)^x \quad \text{for } x = 1, 2, 3, \ldots$$

show that $E(2^X)$ does not exist. This is the famous **Petersburg paradox**, according to which a player's expectation is infinite (does not exist) if he or she is to receive 2^x dollars when, in a series of flips of a balanced coin, the first head appears on the xth flip.

3 Moments

In statistics, the mathematical expectations defined here and in Definition 4, called the **moments** of the distribution of a random variable or simply the **moments** of a random variable, are of special importance.

DEFINITION 2. MOMENTS ABOUT THE ORIGIN. *The* **r*th moment about the origin*** *of a random variable* X*, denoted by* μ_r'*, is the expected value of* X^r*; symbolically*

$$\mu_r' = E(X^r) = \sum_x x^r \cdot f(x)$$

for $r = 0, 1, 2, \ldots$ *when* X *is discrete, and*

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x)dx$$

when X *is continuous.*

It is of interest to note that the term "moment" comes from the field of physics: If the quantities $f(x)$ in the discrete case were point masses acting perpendicularly to the x-axis at distances x from the origin, μ_1' would be the x-coordinate of the center of gravity, that is, the first moment divided by $\sum f(x) = 1$, and μ_2' would be the moment of inertia. This also explains why the moments μ_r' are called moments about the origin: In the analogy to physics, the length of the lever arm is in each case the distance from the origin. The analogy applies also in the continuous case, where μ_1' and μ_2' might be the x-coordinate of the center of gravity and the moment of inertia of a rod of variable density.

When $r = 0$, we have $\mu_0' = E(X^0) = E(1) = 1$ by Corollary 2 of Theorem 2. When $r = 1$, we have $\mu_1' = E(X)$, which is just the expected value of the random variable X, and in view of its importance in statistics we give it a special symbol and a special name.

DEFINITION 3. MEAN OF A DISTRIBUTION. μ_1' *is called the* ***mean*** *of the distribution of* X*, or simply the* ***mean of* X**, *and it is denoted simply by* μ.

The special moments we shall define next are of importance in statistics because they serve to describe the shape of the distribution of a random variable, that is, the shape of the graph of its probability distribution or probability density.

DEFINITION 4. MOMENTS ABOUT THE MEAN. *The* **r*th moment about the mean*** *of a random variable* X*, denoted by* μ_r*, is the expected value of* $(X - \mu)^r$*, symbolically*

$$\mu_r = E\left[(X-\mu)^r\right] = \sum_x (x-\mu)^r \cdot f(x)$$

for $r = 0, 1, 2, \ldots,$ *when* X *is discrete, and*

$$\mu_r = E\left[(X-\mu)^r\right] = \int_{-\infty}^{\infty} (x-u)^r \cdot f(x)dx$$

when X *is continuous.*

Note that $\mu_0 = 1$ and $\mu_1 = 0$ for any random variable for which μ exists (see Exercise 17).

The second moment about the mean is of special importance in statistics because it is indicative of the spread or dispersion of the distribution of a random variable; thus, it is given a special symbol and a special name.

DEFINITION 5. VARIANCE. *μ_2 is called the **variance** of the distribution of X, or simply the **variance of X**, and it is denoted by σ^2, $\sigma_x{}^2$, var(X), or V(X). The positive square root of the variance, σ, is called the **standard deviation of X**.*

Figure 1 shows how the variance reflects the spread or dispersion of the distribution of a random variable. Here we show the histograms of the probability distributions of four random variables with the same mean $\mu = 5$ but variances equaling 5.26, 3.18, 1.66, and 0.88. As can be seen, a small value of σ^2 suggests that we are likely to get a value close to the mean, and a large value of σ^2 suggests that there is a greater probability of getting a value that is not close to the mean. This will be discussed further in Section 4. A brief discussion of how μ_3, the third moment about the mean, describes the **symmetry** or **skewness** (lack of symmetry) of a distribution is given in Exercise 26.

In many instances, moments about the mean are obtained by first calculating moments about the origin and then expressing the μ_r in terms of the μ_r'. To serve this purpose, the reader will be asked to verify a general formula in Exercise 25. Here, let us merely derive the following computing formula for σ^2.

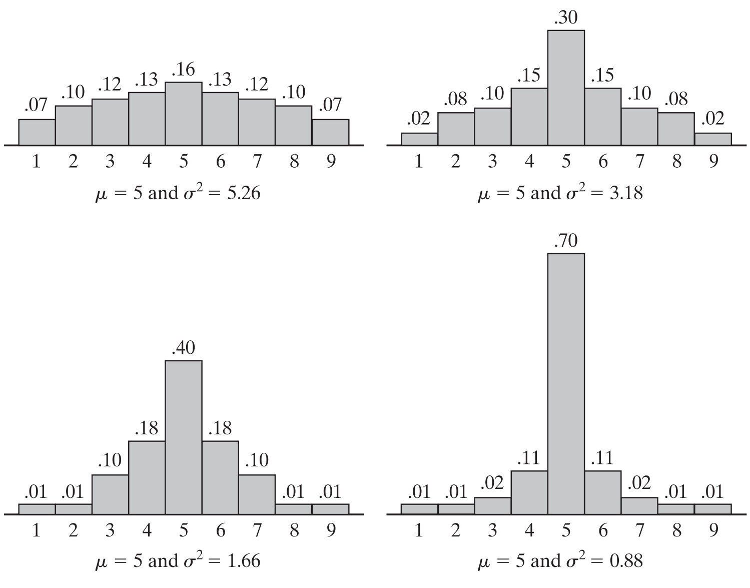

Figure 1. Distributions with different dispersions.

THEOREM 6.

$$\sigma^2 = \mu_2' - \mu^2$$

Proof

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \\ &= \mu_2' - \mu^2\end{aligned}$$

EXAMPLE 10

Use Theorem 6 to calculate the variance of X, representing the number of points rolled with a balanced die.

Solution
First we compute

$$\begin{aligned}\mu = E(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2}\end{aligned}$$

Now,

$$\begin{aligned}\mu_2' = E(X^2) &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\ &= \frac{91}{6}\end{aligned}$$

and it follows that

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

EXAMPLE 11

With reference to Example 2, find the standard deviation of the random variable X.

Solution
In Example 2 we showed that $\mu = E(X) = 0.4413$. Now

$$\begin{aligned}\mu_2' = E(X^2) &= \frac{4}{\pi}\int_0^1 \frac{x^2}{1+x^2}\,dx \\ &= \frac{4}{\pi}\int_0^1 \left(1 - \frac{1}{1+x^2}\right) dx\end{aligned}$$

$$= \frac{4}{\pi} - 1$$
$$= 0.2732$$

and it follows that

$$\sigma^2 = 0.2732 - (0.4413)^2 = 0.0785$$

and $\sigma = \sqrt{0.0785} = 0.2802$.

The following is another theorem that is of importance in work connected with standard deviations or variances.

THEOREM 7. If X has the variance σ^2, then

$$\text{var}(aX + b) = a^2\sigma^2$$

The proof of this theorem will be left to the reader, but let us point out the following corollaries: For $a = 1$, we find that the addition of a constant to the values of a random variable, resulting in a shift of all the values of X to the left or to the right, in no way affects the spread of its distribution; for $b = 0$, we find that if the values of a random variable are multiplied by a constant, the variance is multiplied by the square of that constant, resulting in a corresponding change in the spread of the distribution.

4 Chebyshev's Theorem

To demonstrate how σ or σ^2 is indicative of the spread or dispersion of the distribution of a random variable, let us now prove the following theorem, called **Chebyshev's theorem** after the nineteenth-century Russian mathematician P. L. Chebyshev. We shall prove it here only for the continuous case, leaving the discrete case as an exercise.

THEOREM 8. (*Chebyshev's Theorem*) If μ and σ are the mean and the standard deviation of a random variable X, then for any positive constant k the probability is *at least* $1 - \frac{1}{k^2}$ that X will take on a value within k standard deviations of the mean; symbolically,

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \quad \sigma \neq 0$$

Proof According to Definitions 4 and 5, we write

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx$$

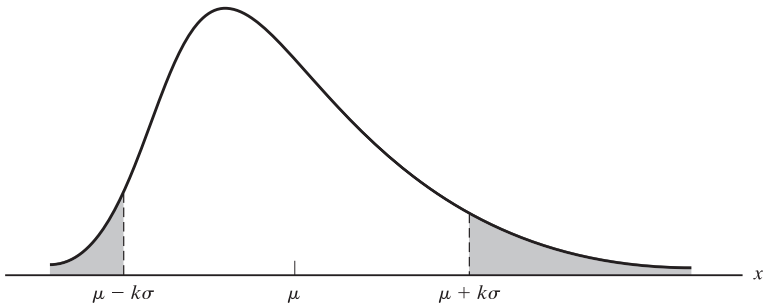

Figure 2. Diagram for proof of Chebyshev's theorem.

Then, dividing the integral into three parts as shown in Figure 2, we get

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

Since the integrand $(x-\mu)^2 \cdot f(x)$ is nonnegative, we can form the inequality

$$\sigma^2 \geqq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

by deleting the second integral. Therefore, since $(x-\mu)^2 \geqq k^2\sigma^2$ for $x \leqq \mu - k\sigma$ or $x \geqq \mu + k\sigma$ it follows that

$$\sigma^2 \geqq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 \cdot f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 \cdot f(x)\,dx$$

and hence that

$$\frac{1}{k^2} \geqq \int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx$$

provided $\sigma^2 \neq 0$. Since the sum of the two integrals on the right-hand side is the probability that X will take on a value less than or equal to $\mu - k\sigma$ or greater than or equal to $\mu + k\sigma$, we have thus shown that

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

and it follows that

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

For instance, the probability is at least $1 - \frac{1}{2^2} = \frac{3}{4}$ that a random variable X will take on a value within two standard deviations of the mean, the probability is at least $1 - \frac{1}{3^2} = \frac{8}{9}$ that it will take on a value within three standard deviations of the mean, and the probability is at least $1 - \frac{1}{5^2} = \frac{24}{25}$ that it will take on a value within five standard deviations of the mean. It is in this sense that σ controls the spread or dispersion of the distribution of a random variable. Clearly, the probability given by Chebyshev's theorem is only a lower bound; whether the probability that a given random variable will take on a value within k standard deviations of the mean is actually greater than $1 - \frac{1}{k^2}$ and, if so, by how much we cannot say, but Chebyshev's theorem assures us that this probability cannot be less than $1 - \frac{1}{k^2}$. Only when the distribution of a random variable is known can we calculate the exact probability.

EXAMPLE 12

If the probability density of X is given by

$$f(x) = \begin{cases} 630x^4(1-x)^4 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability that it will take on a value within two standard deviations of the mean and compare this probability with the lower bound provided by Chebyshev's theorem.

Solution

Straightforward integration shows that $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{44}$, so that $\sigma = \sqrt{1/44}$ or approximately 0.15. Thus, the probability that X will take on a value within two standard deviations of the mean is the probability that it will take on a value between 0.20 and 0.80, that is,

$$\begin{aligned} P(0.20 < X < 0.80) &= \int_{0.20}^{0.80} 630x^4(1-x)^4\,dx \\ &= 0.96 \end{aligned}$$

Observe that the statement "the probability is 0.96" is a much stronger statement than "the probability is at least 0.75," which is provided by Chebyshev's theorem.

---

5 Moment-Generating Functions

Although the moments of most distributions can be determined directly by evaluating the necessary integrals or sums, an alternative procedure sometimes provides considerable simplifications. This technique utilizes **moment-generating functions**.

DEFINITION 6. MOMENT GENERATING FUNCTION. *The* ***moment generating function*** *of a random variable* X*, where it exists, is given by*

$$M_X(t) = E(e^{tX}) = \sum_x e^{tX} \cdot f(x)$$

when X *is discrete, and*

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx$$

when X *is continuous.*

The independent variable is t, and we are usually interested in values of t in the neighborhood of 0.

To explain why we refer to this function as a "moment-generating" function, let us substitute for e^{tx} its Maclaurin's series expansion, that is,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^rx^r}{r!} + \cdots$$

For the discrete case, we thus get

$$\begin{aligned}
M_X(t) &= \sum_x \left[1 + tx + \frac{t^2x^2}{2!} + \cdots + \frac{t^rx^r}{r!} + \cdots\right] f(x) \\
&= \sum_x f(x) + t \cdot \sum_x x f(x) + \frac{t^2}{2!} \cdot \sum_x x^2 f(x) + \cdots + \frac{t^r}{r!} \cdot \sum_x x^r f(x) + \cdots \\
&= 1 + \mu \cdot t + \mu_2' \cdot \frac{t^2}{2!} + \cdots + \mu_r' \cdot \frac{t^r}{r!} + \cdots
\end{aligned}$$

and it can be seen that in the Maclaurin's series of the moment-generating function of X the coefficient of $\frac{t^r}{r!}$ is μ_r', the rth moment about the origin. In the continuous case, the argument is the same.

EXAMPLE 13

Find the moment-generating function of the random variable whose probability density is given by

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and use it to find an expression for μ_r'.

Solution

By definition

$$\begin{aligned}
M_X(t) = E(e^{tX}) &= \int_0^\infty e^{tx} \cdot e^{-x}\, dx \\
&= \int_0^\infty e^{-x(1-t)}\, dx \\
&= \frac{1}{1-t} \quad \text{for } t < 1
\end{aligned}$$

As is well known, when $|t| < 1$ the Maclaurin's series for this moment-generating function is

$$\begin{aligned}
M_X(t) &= 1 + t + t^2 + t^3 + \cdots + t^r + \cdots \\
&= 1 + 1! \cdot \frac{t}{1!} + 2! \cdot \frac{t^2}{2!} + 3! \cdot \frac{t^3}{3!} + \cdots + r! \cdot \frac{t^r}{r!} + \cdots
\end{aligned}$$

and hence $\mu_r' = r!$ for $r = 0, 1, 2, \ldots$.

The main difficulty in using the Maclaurin's series of a moment-generating function to determine the moments of a random variable is usually *not* that of finding the moment-generating function, but that of expanding it into a Maclaurin's series. If we are interested only in the first few moments of a random variable, say, μ_1' and μ_2', their determination can usually be simplified by using the following theorem.

THEOREM 9.
$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'$$

This follows from the fact that if a function is expanded as a power series in t, the coefficient of $\frac{t^r}{r!}$ is the rth derivative of the function with respect to t at $t = 0$.

EXAMPLE 14

Given that X has the probability distribution $f(x) = \frac{1}{8}\binom{3}{x}$ for $x = 0, 1, 2$, and 3, find the moment-generating function of this random variable and use it to determine μ_1' and μ_2'.

Solution
In accordance with Definition 6,

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \frac{1}{8}\cdot\sum_{x=0}^{3} e^{tx}\binom{3}{x} \\ &= \frac{1}{8}(1+3e^t+3e^{2t}+e^{3t}) \\ &= \frac{1}{8}(1+e^t)^3 \end{aligned}$$

Then, by Theorem 9,

$$\mu_1' = M_X'(0) = \left.\frac{3}{8}(1+e^t)^2 e^t\right|_{t=0} = \frac{3}{2}$$

and

$$\mu_2' = M_X''(0) = \left.\frac{3}{4}(1+e^t)e^{2t} + \frac{3}{8}(1+e^t)^2 e^t\right|_{t=0} = 3$$

Often the work involved in using moment-generating functions can be simplified by making use of the following theorem.

THEOREM 10. If a and b are constants, then

1. $M_{X+a}(t) = E[e^{(X+a)t}] = e^{at}\cdot M_X(t)$;
2. $M_{bX}(t) = E(e^{bXt}) = M_X(bt)$;
3. $M_{\frac{X+a}{b}}(t) = E[e^{\left(\frac{X+a}{b}\right)t}] = e^{\frac{a}{b}t}\cdot M_X\left(\frac{t}{b}\right)$.

The proof of this theorem is left to the reader in Exercise 39. The first part of the theorem is of special importance when $a = -\mu$, and the third part is of special importance when $a = -\mu$ and $b = \sigma$, in which case

$$M_{\frac{X-\mu}{\sigma}}(t) = e^{-\frac{\mu t}{\sigma}} \cdot M_X\left(\frac{t}{\sigma}\right)$$

Exercises

17. With reference to Definition 4, show that $\mu_0 = 1$ and that $\mu_1 = 0$ for any random variable for which $E(X)$ exists.

18. Find μ, μ_2', and σ^2 for the random variable X that has the probability distribution $f(x) = \frac{1}{2}$ for $x = -2$ and $x = 2$.

19. Find μ, μ_2', and σ^2 for the random variable X that has the probability density

$$f(x) = \begin{cases} \dfrac{x}{2} & \text{for } 0 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

20. Find μ_r' and σ^2 for the random variable X that has the probability density

$$f(x) = \begin{cases} \dfrac{1}{\ln 3} \cdot \dfrac{1}{x} & \text{for } 1 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

21. Prove Theorem 7.

22. With reference to Exercise 8, find the variance of $g(X) = 2X + 3$.

23. If the random variable X has the mean μ and the standard deviation σ, show that the random variable Z whose values are related to those of X by means of the equation $z = \frac{x-\mu}{\sigma}$ has

$$E(Z) = 0 \quad \text{and} \quad \text{var}(Z) = 1$$

A distribution that has the mean 0 and the variance 1 is said to be in **standard form**, and when we perform the above change of variable, we are said to be **standardizing** the distribution of X.

24. If the probability density of X is given by

$$f(x) = \begin{cases} 2x^{-3} & \text{for } x > 1 \\ 0 & \text{elsewhere} \end{cases}$$

check whether its mean and its variance exist.

25. Show that

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}' \cdot \mu + \cdots + (-1)^i \binom{r}{i}\mu_{r-i}' \cdot \mu^i$$
$$+ \cdots + (-1)^{r-1}(r-1) \cdot \mu^r$$

for $r = 1, 2, 3, \ldots$, and use this formula to express μ_3 and μ_4 in terms of moments about the origin.

26. The **symmetry** or **skewness** (lack of symmetry) of a distribution is often measured by means of the quantity

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

Use the formula for μ_3 obtained in Exercise 25 to determine α_3 for each of the following distributions (which have equal means and standard deviations):

(a) $f(1) = 0.05, f(2) = 0.15, f(3) = 0.30, f(4) = 0.30, f(5) = 0.15$, and $f(6) = 0.05$;

(b) $f(1) = 0.05, f(2) = 0.20, f(3) = 0.15, f(4) = 0.45, f(5) = 0.10$, and $f(6) = 0.05$.

Also draw histograms of the two distributions and note that whereas the first is symmetrical, the second has a "tail" on the left-hand side and is said to be **negatively skewed**.

27. The extent to which a distribution is peaked or flat, also called the **kurtosis** of the distribution, is often measured by means of the quantity

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

Use the formula for μ_4 obtained in Exercise 25 to find α_4 for each of the following symmetrical distributions, of which the first is more peaked (narrow humped) than the second:

(a) $f(-3) = 0.06, f(-2) = 0.09, f(-1) = 0.10, f(0) = 0.50, f(1) = 0.10, f(2) = 0.09$, and $f(3) = 0.06$;

(b) $f(-3) = 0.04, f(-2) = 0.11, f(-1) = 0.20, f(0) = 0.30, f(1) = 0.20, f(2) = 0.11$, and $f(3) = 0.04$.

28. Duplicate the steps used in the proof of Theorem 8 to prove Chebyshev's theorem for a discrete random variable X.

29. Show that if X is a random variable with the mean μ for which $f(x) = 0$ for $x < 0$, then for any positive constant a,

$$P(X \geqq a) \leqq \frac{\mu}{a}$$

This inequality is called **Markov's inequality**, and we have given it here mainly because it leads to a relatively simple alternative proof of Chebyshev's theorem.

30. Use the inequality of Exercise 29 to prove Chebyshev's theorem. [*Hint*: Substitute $(X - \mu)^2$ for X.]

31. What is the smallest value of k in Chebyshev's theorem for which the probability that a random variable will take on a value between $\mu - k\sigma$ and $\mu + k\sigma$ is
(a) at least 0.95;
(b) at least 0.99?

32. If we let $k\sigma = c$ in Chebyshev's theorem, what does this theorem assert about the probability that a random variable will take on a value between $\mu - c$ and $\mu + c$?

33. Find the moment-generating function of the continuous random variable X whose probability density is given by

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and use it to find μ_1', μ_2', and σ^2.

34. Find the moment-generating function of the discrete random variable X that has the probability distribution

$$f(x) = 2\left(\frac{1}{3}\right)^x \quad \text{for } x = 1, 2, 3, \ldots$$

and use it to determine the values of μ_1' and μ_2'.

35. If we let $R_X(t) = \ln M_X(t)$, show that $R_X'(0) = \mu$ and $R_X''(0) = \sigma^2$. Also, use these results to find the mean and the variance of a random variable X having the moment-generating function

$$M_X(t) = e^{4(e^t - 1)}$$

36. Explain why there can be no random variable for which $M_X(t) = \frac{t}{1-t}$.

37. Show that if a random variable has the probability density

$$f(x) = \frac{1}{2}\,e^{-|x|} \quad \text{for } -\infty < x < \infty$$

its moment-generating function is given by

$$M_X(t) = \frac{1}{1 - t^2}$$

38. With reference to Exercise 37, find the variance of the random variable by
(a) expanding the moment-generating function as an infinite series and reading off the necessary coefficients;
(b) using Theorem 9.

39. Prove the three parts of Theorem 10.

40. Given the moment-generating function $M_X(t) = e^{3t+8t^2}$, find the moment-generating function of the random variable $Z = \frac{1}{4}(X - 3)$, and use it to determine the mean and the variance of Z.

6 Product Moments

To continue the discussion of Section 3, let us now present the **product moments** of two random variables.

DEFINITION 7. PRODUCT MOMENTS ABOUT THE ORIGIN. *The* **r*th* *and* s*th product moment about the origin*** *of the random variables* X *and* Y*, denoted by* $\mu'_{r,s}$*, is the expected value of* X^rY^s*; symbolically,*

$$\mu'_{r,s} = E(X^rY^s) = \sum_x \sum_y x^r y^s \cdot f(x, y)$$

for $r = 0, 1, 2, \ldots$ *and* $s = 0, 1, 2, \ldots$ *when* X *and* Y *are discrete, and*

$$\mu'_{r,s} = E(X^rY^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s \cdot f(x, y)\,dx\,dy$$

when X *and* Y *are continuous.*

In the discrete case, the double summation extends over the entire joint range of the two random variables. Note that $\mu'_{1,0} = E(X)$, which we denote here by μ_X, and that $\mu'_{0,1} = E(Y)$, which we denote here by μ_Y.

Analogous to Definition 4, let us now state the following definition of product moments about the respective means.

DEFINITION 8. PRODUCT MOMENTS ABOUT THE MEAN. *The* r*th and* s*th* ***product moment about the means*** *of the random variables* X *and* Y*, denoted by* $\mu_{r,s}$*, is the expected value of* $(X - \mu X)^r (Y - \mu_Y)^s$*; symbolically,*

$$\mu_{r,s} = E[(X - \mu_X)^r (Y - \mu_Y)^s]$$
$$= \sum_x \sum_y (x - \mu_X)^r (y - \mu_Y)^s \cdot f(x, y)$$

for $r = 0, 1, 2, \ldots$ *and* $s = 0, 1, 2, \ldots$ *when* X *and* Y *are discrete, and*

$$\mu_{r,s} = E[(X - \mu_X)^r (Y - \mu_Y)^s]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s \cdot f(x, y) dx dy$$

when X *and* Y *are continuous.*

In statistics, $\mu_{1,1}$ is of special importance because it is indicative of the relationship, if any, between the values of X and Y; thus, it is given a special symbol and a special name.

DEFINITION 9. COVARIANCE. $\mu_{1,1}$ *is called the* ***covariance*** *of* X *and* Y*, and it is denoted by* σ_{XY}*,* $cov(X, Y)$*, or* $C(X, Y)$.

Observe that if there is a high probability that large values of X will go with large values of Y and small values of X with small values of Y, the covariance will be positive; if there is a high probability that large values of X will go with small values of Y, and vice versa, the covariance will be negative. It is in this sense that the covariance measures the relationship, or association, between the values of X and Y.

Let us now prove the following result, analogous to Theorem 6, which is useful in actually determining covariances.

THEOREM 11.

$$\sigma_{XY} = \mu'_{1,\,1} - \mu_X \mu_Y$$

Proof Using the various theorems about expected values, we can write

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$$
$$= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y$$
$$= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y$$
$$= \mu'_{1,\,1} - \mu_X\mu_Y$$

EXAMPLE 15

The joint and marginal probabilities of X and Y, the numbers of aspirin and sedative caplets among two caplets drawn at random from a bottle containing three aspirin, two sedative, and four laxative caplets, are recorded as follows:

$$
\begin{array}{cc|ccc|c}
 & & & x & & \\
 & & 0 & 1 & 2 & \\
\hline
 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{12} & \frac{7}{12} \\
y & 1 & \frac{2}{9} & \frac{1}{6} & & \frac{7}{18} \\
 & 2 & \frac{1}{36} & & & \frac{1}{36} \\
\hline
 & & \frac{5}{12} & \frac{1}{2} & \frac{1}{12} &
\end{array}
$$

Find the covariance of X and Y.

Solution

Referring to the joint probabilities given here, we get

$$
\begin{aligned}
\mu'_{1,\,1} &= E(XY) \\
&= 0\cdot 0\cdot \frac{1}{6} + 0\cdot 1\cdot \frac{2}{9} + 0\cdot 2\cdot \frac{1}{36} + 1\cdot 0\cdot \frac{1}{3} + 1\cdot 1\cdot \frac{1}{6} + 2\cdot 0\cdot \frac{1}{12} \\
&= \frac{1}{6}
\end{aligned}
$$

and using the marginal probabilities, we get

$$\mu_X = E(X) = 0\cdot \frac{5}{12} + 1\cdot \frac{1}{2} + 2\cdot \frac{1}{12} = \frac{2}{3}$$

and

$$\mu_Y = E(Y) = 0\cdot \frac{7}{12} + 1\cdot \frac{7}{18} + 2\cdot \frac{1}{36} = \frac{4}{9}$$

It follows that

$$\sigma_{XY} = \frac{1}{6} - \frac{2}{3}\cdot \frac{4}{9} = -\frac{7}{54}$$

The negative result suggests that the more aspirin tablets we get the fewer sedative tablets we will get, and vice versa, and this, of course, makes sense.

EXAMPLE 16

Find the covariance of the random variables whose joint probability density is given by

$$f(x, y) = \begin{cases} 2 & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Solution
Evaluating the necessary integrals, we get

$$\mu_X = \int_0^1 \int_0^{1-x} 2x\, dy\, dx = \frac{1}{3}$$
$$\mu_Y = \int_0^1 \int_0^{1-x} 2y\, dy\, dx = \frac{1}{3}$$

and

$$\sigma'_{1,1} = \int_0^1 \int_0^{1-x} 2xy\, dy\, dx = \frac{1}{12}$$

It follows that

$$\sigma_{XY} = \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{36}$$

As far as the relationship between X and Y is concerned, observe that if X and Y are independent, their covariance is zero; symbolically, we have the following theorem.

Theorem 12. If X and Y are independent, then $E(XY) = E(X) \cdot E(Y)$ and $\sigma_{XY} = 0$.

Proof For the discrete case we have, by definition,

$$E(XY) = \sum_x \sum_y xy \cdot f(x, y)$$

Since X and Y are independent, we can write $f(x, y) = g(x) \cdot h(y)$, where $g(x)$ and $h(y)$ are the values of the marginal distributions of X and Y, and we get

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy \cdot g(x)h(y) \\ &= \left[\sum_x x \cdot g(x)\right]\left[\sum_y y \cdot h(y)\right] \\ &= E(X) \cdot E(Y) \end{aligned}$$

Hence,

$$\begin{aligned} \sigma_{XY} &= \mu'_{1,1} - \mu_X \mu_Y \\ &= E(X) \cdot E(Y) - E(X) \cdot E(Y) \\ &= 0 \end{aligned}$$

It is of interest to note that the independence of two random variables implies a zero covariance, but a zero covariance does not necessarily imply their independence. This is illustrated by the following example (see also Exercises 46 and 47).

EXAMPLE 17

If the joint probability distribution of X and Y is given by

$$
\begin{array}{cc|c|c|c|c}
 & & & x & & \\
 & & -1 & 0 & 1 & \\
\hline
 & -1 & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{2}{3} \\
\hline
y & 0 & 0 & 0 & 0 & 0 \\
\hline
 & 1 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{3} \\
\hline
 & & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} &
\end{array}
$$

show that their covariance is zero even though the two random variables are not independent.

Solution

Using the probabilities shown in the margins, we get

$$
\begin{aligned}
\mu_X &= (-1)\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0 \\
\mu_Y &= (-1)\cdot\frac{2}{3} + 0\cdot 0 + 1\cdot\frac{1}{3} = -\frac{1}{3}
\end{aligned}
$$

and

$$
\begin{aligned}
\mu'_{1,1} &= (-1)(-1)\cdot\frac{1}{6} + 0(-1)\cdot\frac{1}{3} + 1(-1)\cdot\frac{1}{6} + (-1)1\cdot\frac{1}{6} + 1\cdot 1\cdot\frac{1}{6} \\
&= 0
\end{aligned}
$$

Thus, $\sigma_{XY} = 0 - 0(-\frac{1}{3}) = 0$, the covariance is zero, but the two random variables are not independent. For instance, $f(x,y) \neq g(x)\cdot h(y)$ for $x = -1$ and $y = -1$.

Product moments can also be defined for the case where there are more than two random variables. Here let us merely state the important result, in the following theorem.

> **THEOREM 13.** If $X_1, X_2, \ldots, X_n$ are independent, then
>
> $$E(X_1 X_2 \cdot \ldots \cdot X_n) = E(X_1)\cdot E(X_2)\cdot \ldots \cdot E(X_n)$$

This is a generalization of the first part of Theorem 12.

7 Moments of Linear Combinations of Random Variables

In this section we shall derive expressions for the mean and the variance of a linear combination of n random variables and the covariance of two linear combinations of n random variables. Applications of these results will be important in our later discussion of sampling theory and problems of statistical inference.

THEOREM 14. If $X_1, X_2, \ldots, X_n$ are random variables and

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

where $a_1, a_2, \ldots, a_n$ are constants, then

$$E(Y) = \sum_{i=1}^{n} a_i E(X_i)$$

and

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i) + 2 \sum_{i<j}\sum a_i a_j \cdot \text{cov}(X_i X_j)$$

where the double summation extends over all values of i and j, from 1 to n, for which $i < j$.

Proof From Theorem 5 with $g_i(X_1, X_2, \ldots, X_k) = X_i$ for $i = 0, 1, 2, \ldots, n$, it follows immediately that

$$E(Y) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

and this proves the first part of the theorem. To obtain the expression for the variance of Y, let us write μ_i for $E(X_i)$ so that we get

$$\text{var}(Y) = E\left([Y - E(Y)]^2\right) = E\left\{\left[\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i)\right]^2\right\}$$

$$= E\left\{\left[\sum_{i=1}^{n} a_i (X_i - \mu_i)\right]^2\right\}$$

Then, expanding by means of the multinomial theorem, according to which $(a+b+c+d)^2$, for example, equals $a^2+b^2+c^2+d^2+2ab+2ac+2ad+2bc+2bd+2cd$, and again referring to Theorem 5, we get

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 E[(X_i - \mu_i)^2] + 2 \sum_{i<j}\sum a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)]$$

$$= \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i) + 2 \sum_{i<j}\sum a_i a_j \cdot \text{cov}(X_i, X_j)$$

Note that we have tacitly made use of the fact that $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$.

Since $\mathrm{cov}(X_i, X_j) = 0$ when X_i and X_j are independent, we obtain the following corollary.

COROLLARY 3. If the random variables $X_1, X_2, \ldots, X_n$ are independent and $Y = \sum_{i=1}^{n} a_i X_i$, then

$$\mathrm{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \mathrm{var}(X_i)$$

EXAMPLE 18

If the random variables X, Y, and Z have the means $\mu_X = 2, \mu_Y = -3$, and $\mu_Z = 4$, the variances $\sigma_X^2 = 1, \sigma_Y^2 = 5$, and $\sigma_Z^2 = 2$, and the covariances $\mathrm{cov}(X,Y) = -2$, $\mathrm{cov}(X,Z) = -1$, and $\mathrm{cov}(Y,Z) = 1$, find the mean and the variance of $W = 3X - Y + 2Z$.

Solution

By Theorem 14, we get

$$\begin{aligned} E(W) &= E(3X - Y + 2Z) \\ &= 3E(X) - E(Y) + 2E(Z) \\ &= 3 \cdot 2 - (-3) + 2 \cdot 4 \\ &= 17 \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}(W) &= 9\,\mathrm{var}(X) + \mathrm{var}(Y) + 4\,\mathrm{var}(Z) - 6\,\mathrm{cov}(X,Y) \\ &\quad + 12\,\mathrm{cov}(X,Z) - 4\,\mathrm{cov}(Y,Z) \\ &= 9 \cdot 1 + 5 + 4 \cdot 2 - 6(-2) + 12(-1) - 4 \cdot 1 \\ &= 18 \end{aligned}$$

The following is another important theorem about linear combinations of random variables; it concerns the covariance of two linear combinations of n random variables.

THEOREM 15. If $X_1, X_2, \ldots, X_n$ are random variables and

$$Y_1 = \sum_{i=1}^{n} a_i X_i \quad \text{and} \quad Y_2 = \sum_{i=1}^{n} b_i X_i$$

where $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are constants, then

$$\mathrm{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \mathrm{var}(X_i) + \sum\sum_{i<j} (a_i b_j + a_j b_i) \cdot \mathrm{cov}(X_i, X_j)$$

The proof of this theorem, which is very similar to that of Theorem 14, will be left to the reader in Exercise 52.

Since $\mathrm{cov}(X_i, X_j) = 0$ when X_i and X_j are independent, we obtain the following corollary.

COROLLARY 4. If the random variables $X_1, X_2, \ldots, X_n$ are independent, $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$, then

$$\operatorname{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \operatorname{var}(X_i)$$

EXAMPLE 19

If the random variables X, Y, and Z have the means $\mu_X = 3, \mu_Y = 5$, and $\mu_Z = 2$, the variances $\sigma_X^2 = 8, \sigma_Y^2 = 12$, and $\sigma_Z^2 = 18$, and $\operatorname{cov}(X, Y) = 1, \operatorname{cov}(X, Z) = -3$, and $\operatorname{cov}(Y, Z) = 2$, find the covariance of

$$U = X + 4Y + 2Z \quad \text{and} \quad V = 3X - Y - Z$$

Solution

By Theorem 15, we get

$$\begin{aligned}
\operatorname{cov}(U, V) &= \operatorname{cov}(X + 4Y + 2Z, 3X - Y - Z) \\
&= 3 \operatorname{var}(X) - 4 \operatorname{var}(Y) - 2 \operatorname{var}(Z) + 11 \operatorname{cov}(X, Y) \\
&\quad + 5 \operatorname{cov}(X, Z) - 6 \operatorname{cov}(Y, Z) \\
&= 3 \cdot 8 - 4 \cdot 12 - 2 \cdot 18 + 11 \cdot 1 + 5(-3) - 6 \cdot 2 \\
&= -76
\end{aligned}$$

8 Conditional Expectations

Conditional probabilities are obtained by adding the values of conditional probability distributions, or integrating the values of conditional probability densities. **Conditional expectations** of random variables are likewise defined in terms of their conditional distributions.

DEFINITION 10. CONDITIONAL EXPECTATION. *If* X *is a discrete random variable, and* f(x|y) *is the value of the conditional probability distribution of* X *given* Y = y *at* x*, the* ***conditional expectation of* u(X) *given* Y = y** *is*

$$E[u(X)|y)] = \sum_{x} u(x) \cdot f(x|y)$$

Correspondingly, if X *is a continuous variable and* f(x|y) *is the value of the conditional probability distribution of* X *given* Y = y *at* x*, the* ***conditional expectation of* u(X) *given* Y = y** *is*

$$E[(u(X)|y)] = \int_{=\infty}^{\infty} u(x) \cdot f(x|y) dx$$

Similar expressions based on the conditional probability distribution or density of Y given $X = x$ define the conditional expectation of $\upsilon(Y)$ given $X = x$.

If we let $u(X) = X$ in Definition 10, we obtain the **conditional mean** of the random variable X given $Y = y$, which we denote by

$$\mu_{X|y} = E(X|y)$$

Correspondingly, the **conditional variance** of X given $Y = y$ is

$$\begin{aligned}\sigma^2_{X|y} &= E[(X - \mu_{X|y})^2|y] \\ &= E(X^2|y) - \mu^2_{X|y}\end{aligned}$$

where $E(X^2|y)$ is given by Definition 10 with $u(X) = X^2$. The reader should not find it difficult to generalize Definition 10 for conditional expectations involving more than two random variables.

EXAMPLE 20

If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the conditional mean and the conditional variance of X given $Y = \frac{1}{2}$.

Solution

For these random variables the conditional density of X given $Y = y$ is

$$f(x|y) = \begin{cases} \dfrac{2x+4y}{1+4y} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

so that

$$f\left(x\middle|\frac{1}{2}\right) = \begin{cases} \frac{2}{3}(x+1) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Thus, $\mu_{X|\frac{1}{2}}$ is given by

$$\begin{aligned}E\left(X\middle|\frac{1}{2}\right) &= \int_0^1 \frac{2}{3}x(x+1)\,dx \\ &= \frac{5}{9}\end{aligned}$$

Next we find

$$\begin{aligned}E\left(X^2\middle|\frac{1}{2}\right) &= \int_0^1 \frac{2}{3}x^2(x+1)\,dx \\ &= \frac{7}{18}\end{aligned}$$

and it follows that

$$\sigma^2_{X|\frac{1}{2}} = \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162}$$

Exercises

41. Prove that $\text{cov}(X, Y) = \text{cov}(Y, X)$ for both discrete and continuous random variables X and Y.

42. If X and Y have the joint probability distribution $f(x, y) = \frac{1}{4}$ for $x = -3$ and $y = -5$, $x = -1$ and $y = -1$, $x = 1$ and $y = 1$, and $x = 3$ and $y = 5$, find $\text{cov}(X, Y)$.

43. This has been intentionally omitted for this edition.

44. This has been intentionally omitted for this edition.

45. This has been intentionally omitted for this edition.

46. If X and Y have the joint probability distribution $f(-1, 0) = 0$, $f(-1, 1) = \frac{1}{4}$, $f(0, 0) = \frac{1}{6}$, $f(0, 1) = 0$, $f(1, 0) = \frac{1}{12}$, and $f(1, 1) = \frac{1}{2}$, show that

(a) $\text{cov}(X, Y) = 0$;

(b) the two random variables are not independent.

47. If the probability density of X is given by

$$f(x) = \begin{cases} 1+x & \text{for } -1 < x \leqq 0 \\ 1-x & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and $U = X$ and $V = X^2$, show that

(a) $\text{cov}(U, V) = 0$;

(b) U and V are dependent.

48. For k random variables $X_1, X_2, \ldots, X_k$, the values of their **joint moment-generating function** are given by

$$E\left(e^{t_1X_1+t_2X_2+\cdots+t_kX_k}\right)$$

(a) Show for either the discrete case or the continuous case that the partial derivative of the joint moment-generating function with respect to t_i at $t_1 = t_2 = \cdots = t_k = 0$ is $E(X_i)$.

(b) Show for either the discrete case or the continuous case that the second partial derivative of the joint moment-generating function with respect to t_i and t_j, $i \neq j$, at $t_1 = t_2 = \cdots = t_k = 0$ is $E(X_iX_j)$.

(c) If two random variables have the joint density given by

$$f(x, y) = \begin{cases} e^{-x-y} & \text{for } x > 0,\ y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find their joint moment-generating function and use it to determine the values of $E(XY)$, $E(X)$, $E(Y)$, and $\text{cov}(X, Y)$.

49. If X_1, X_2, and X_3 are independent and have the means 4, 9, and 3 and the variances 3, 7, and 5, find the mean and the variance of

(a) $Y = 2X_1 - 3X_2 + 4X_3$;

(b) $Z = X_1 + 2X_2 - X_3$.

50. Repeat both parts of Exercise 49, dropping the assumption of independence and using instead the information that $\text{cov}(X_1, X_2) = 1$, $\text{cov}(X_2, X_3) = -2$, and $\text{cov}(X_1, X_3) = -3$.

51. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{3}(x+y) & \text{for } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the variance of $W = 3X + 4Y - 5$.

52. Prove Theorem 15.

53. Express $\text{var}(X + Y)$, $\text{var}(X - Y)$, and $\text{cov}(X + Y, X - Y)$ in terms of the variances and covariance of X and Y.

54. If $\text{var}(X_1) = 5$, $\text{var}(X_2) = 4$, $\text{var}(X_3) = 7$, $\text{cov}(X_1, X_2) = 3$, $\text{cov}(X_1, X_3) = -2$, and X_2 and X_3 are independent, find the covariance of $Y_1 = X_1 - 2X_2 + 3X_3$ and $Y_2 = -2X_1 + 3X_2 + 4X_3$.

55. With reference to Exercise 49, find $\text{cov}(Y, Z)$.

56. This question has been intentionally omitted for this edition.

57. This question has been intentionally omitted for this edition.

58. This question has been intentionally omitted for this edition.

59. This question has been intentionally omitted for this edition.

60. **(a)** Show that the conditional distribution function of the continuous random variable X, given $a < X \leqq b$, is given by

$$F(x|a < X \leqq b) = \begin{cases} 0 & \text{for } x \leqq a \\ \dfrac{F(x) - F(a)}{F(b) - F(a)} & \text{for } a < x \leqq b \\ 1 & \text{for } x > b \end{cases}$$

(b) Differentiate the result of part (a) with respect to x to find the conditional probability density of X given $a < X \leqq b$, and show that

$$E[u(X)|a < X \leqq b] = \frac{\displaystyle\int_a^b u(x)f(x)\,dx}{\displaystyle\int_a^b f(x)\,dx}$$

9 The Theory in Practice

Empirical distributions, those arising from data, can be described by their shape. We will discuss **descriptive measures**, calculated from data, that extend the methodology of describing data. These descriptive measures are based on the ideas of moments, given in Section 3.

The analog of the first moment, $\mu_1' = \mu$, is the **sample mean**, $\bar{x}$, defined as

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

where $i = 1, 2, \ldots, n$ and n is the number of observations.

The usefulness of the sample mean as a description of data can be envisioned by imagining that the histogram of a data distribution has been cut out of a piece of cardboard and balanced by inserting a fulcrum along the horizontal axis. This balance point corresponds to the mean of the data. Thus, the mean can be thought of as the centroid of the data and, as such, it describes its **location**.

The mean is an excellent measure of location for symmetric or nearly symmetric distributions. But it can be misleading when used to measure the location of highly skewed data. To give an example, suppose, in a small company, the annual salaries of its 10 employees (rounded to the nearest \$1,000) are 25, 18, 36, 28, 16, 20, 29, 32, 41, and 150. The mean of these observations is \$39,500. One of the salaries, namely \$150,000, is much higher than the others (it's what the owner pays himself) and only one other employee earns as much as \$39,500. Suppose the owner, in a recruiting ad, claimed that "Our company pays an average salary of \$39,500." He would be technically correct, but very misleading.

Other descriptive measures for the location of data should be used in cases like the one just described. The **median** describes the center of the data as the middle point of the observations. If the data are ranked from, say, smallest to largest, the median becomes observation number $n/2$ if n is an even integer, and it is defined as the mean value of observations $\frac{(n-1)}{2}$ and $\frac{(n+1)}{2}$ if n is an odd integer. The median of the 10 observations given in the preceding example is \$28,000, and it is a much better description of what an employee of this company can expect to earn. You may very well have heard the term "median income" for, say, the incomes of American families. The median is used instead of the mean here because it is well known that the distribution of family incomes in the United States is highly skewed—the great majority of families earn low to moderate incomes, but a relatively few have very high incomes.

The **dispersion** of data also is important in its description. Give the location of data, one reasonably wants to know how closely the observations are grouped around this value. A reasonable measure of dispersion can be based on the square root of the second moment about the mean, σ. The **sample standard deviation**, s, is calculated analogously to the second moment, as follows:

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

Since this formula requires first the calculation of the mean, then subtraction of the mean from each observation before squaring and adding, it is much easier to use the following *calculating formula* for s:

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

Note that in both formulas we divide by $n-1$ instead of n. Using either formula for the calculation of s requires tedious calculation, but every statistical computer program in common use will calculate both the sample mean and the sample standard deviation once the data have been inputted.

EXAMPLE 21

The following are the lengths (in feet) of 10 steel beams rolled in a steel mill and cut to a nominal length of 12 feet:

11.8 12.1 12.5 11.7 11.9 12.0 12.2 11.5 11.9 12.2

Calculate the mean length and its standard deviation. Is the mean a reasonable measure of the location of the data? Why or why not?

Solution

The mean is given by the sum of the observations, $11.8 + 12.1 + \ldots 12.2 = 119.8$, divided by 10, or $\bar{x} = 11.98$ feet. To calculate the standard deviation, we first calculate the sum of the squares of the observations, $(11.8)^2 + (12.1)^2 + \ldots + (12.2)^2 = 1,435.94$. Then substituting into the formula for s, we obtain $s^2 = (10)(1435.94) - (119.8)^2/(10)(9) = 0.082$ foot. Taking the square root, we obtain $s = 0.29$. The mean, 11.98 feet, seems to be a reasonable measure of location inasmuch as the data seem to be approximately symmetrically distributed.

The standard deviation is not the only measure of the dispersion, or variability of data. The **sample range** sometimes is used for this purpose. To calculate the range, we find the largest and the smallest observations, x_l and x_s, defining the range to be

$$r = x_l - x_s$$

This measure of dispersion is used only for small samples; for larger and larger sample sizes, the range becomes a poorer and poorer measure of dispersion.

Applied Exercises SECS. 1–2

61. This question has been intentionally omitted for this edition.

62. The probability that Ms. Brown will sell a piece of property at a profit of \$3,000 is $\frac{3}{20}$, the probability that she will sell it at a profit of \$1,500 is $\frac{7}{20}$, the probability that she will break even is $\frac{7}{20}$, and the probability that she will lose \$1,500 is $\frac{3}{20}$. What is her expected profit?

63. A game of chance is considered **fair**, or **equitable**, if each player's expectation is equal to zero. If someone pays us \$10 each time that we roll a 3 or a 4 with a balanced die, how much should we pay that person when we roll a 1, 2, 5, or 6 to make the game equitable?

64. The manager of a bakery knows that the number of chocolate cakes he can sell on any given day is a random variable having the probability distribution $f(x) = \frac{1}{6}$ for $x = 0, 1, 2, 3, 4$, and 5. He also knows that there is a profit of \$1.00 for each cake that he sells and a loss (due to spoilage) of \$0.40 for each cake that he does not sell. Assuming that each cake can be sold only on the day it is made, find the baker's expected profit for a day on which he bakes

(a) one of the cakes;

(b) two of the cakes;

(c) three of the cakes;

(d) four of the cakes;

(e) five of the cakes.

How many should he bake in order to maximize his expected profit?

65. If a contractor's profit on a construction job can be looked upon as a continuous random variable having the probability density

$$f(x) = \begin{cases} \frac{1}{18}(x+1) & \text{for } -1 < x < 5 \\ 0 & \text{elsewhere} \end{cases}$$

where the units are in \$1,000, what is her expected profit?

66. This question has been intentionally omitted for this edition.

67. This question has been intentionally omitted for this edition.

68. This question has been intentionally omitted for this edition.

69. Mr. Adams and Ms. Smith are betting on repeated flips of a coin. At the start of the game Mr. Adams has a dollars and Ms. Smith has b dollars, at each flip the loser pays the winner one dollar, and the game continues until either player is "ruined." Making use of the fact that in an equitable game each player's mathematical expectation is zero, find the probability that Mr. Adams will win Ms. Smith's b dollars before he loses his a dollars.

SECS. 3–5

70. With reference to Example 1, find the variance of the number of television sets with white cords.

71. The amount of time it takes a person to be served at a given restaurant is a random variable with the probability density

$$f(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the mean and the variance of this random variable.

72. This question has been intentionally omitted for this edition.

73. This question has been intentionally omitted for this edition.

74. The following are some applications of the Markov inequality of Exercise 29:

(a) The scores that high school juniors get on the verbal part of the PSAT/NMSQT test may be looked upon as values of a random variable with the mean $\mu = 41$. Find an upper bound to the probability that one of the students will get a score of 65 or more.

(b) The weight of certain animals may be looked upon as a random variable with a mean of 212 grams. If none of the animals weighs less than 165 grams, find an upper bound to the probability that such an animal will weigh at least 250 grams.

75. The number of marriage licenses issued in a certain city during the month of June may be looked upon as a random variable with $\mu = 124$ and $\sigma = 7.5$. According to Chebyshev's theorem, with what probability can we assert that between 64 and 184 marriage licenses will be issued there during the month of June?

76. A study of the nutritional value of a certain kind of bread shows that the amount of thiamine (vitamin B_1) in a slice may be looked upon as a random variable with $\mu = 0.260$ milligram and $\sigma = 0.005$ milligram. According to Chebyshev's theorem, between what values must be the thiamine content of

(a) at least $\frac{35}{36}$ of all slices of this bread;

(b) at least $\frac{143}{144}$ of all slices of this bread?

77. With reference to Exercise 71, what can we assert about the amount of time it takes a person to be served at the given restaurant if we use Chebyshev's theorem with $k = 1.5$? What is the corresponding probability rounded to four decimals?

SECS. 6–9

78. A quarter is bent so that the probabilities of heads and tails are 0.40 and 0.60. If it is tossed twice, what is the covariance of Z, the number of heads obtained on the first toss, and W, the total number of heads obtained in the two tosses of the coin?

79. The inside diameter of a cylindrical tube is a random variable with a mean of 3 inches and a standard deviation of 0.02 inch, the thickness of the tube is a random variable with a mean of 0.3 inch and a standard deviation of 0.005 inch, and the two random variables are independent. Find the mean and the standard deviation of the outside diameter of the tube.

80. The length of certain bricks is a random variable with a mean of 8 inches and a standard deviation of 0.1 inch, and the thickness of the mortar between two bricks is a random variable with a mean of 0.5 inch and a standard deviation of 0.03 inch. What is the mean and the standard deviation of the length of a wall made of 50 of these bricks laid side by side, if we can assume that all the random variables involved are independent?

81. If heads is a success when we flip a coin, getting a six is a success when we roll a die, and getting an ace is a success when we draw a card from an ordinary deck of 52 playing cards, find the mean and the standard deviation of the total number of successes when we
(a) flip a balanced coin, roll a balanced die, and then draw a card from a well-shuffled deck;
(b) flip a balanced coin three times, roll a balanced die twice, and then draw a card from a well-shuffled deck.

82. If we alternately flip a balanced coin and a coin that is loaded so that the probability of getting heads is 0.45, what are the mean and the standard deviation of the number of heads that we obtain in 10 flips of these coins?

83. This question has been intentionally omitted for this edition.

84. This question has been intentionally omitted for this edition.

85. The amount of time (in minutes) that an executive of a certain firm talks on the telephone is a random variable having the probability density

$$f(x) = \begin{cases} \dfrac{x}{4} & \text{for } 0 < x \leqq 2 \\ \dfrac{4}{x^3} & \text{for } x > 2 \\ 0 & \text{elsewhere} \end{cases}$$

With reference to part (b) of Exercise 60, find the expected length of one of these telephone conversations that has lasted at least 1 minute.

Answers to Odd-Numbered Exercises

1 **(a)** $g_1 = 0, g_2 = 1, g_3 = 4$, and $g_4 = 9$; **(b)** $f(0), f(-1) + f(1), f(-2) + f(2)$, and $f(3)$; **(c)** $0 \cdot f(0) + 1 \cdot \{f(-1) + f(1)\} + 4 \cdot \{f(-2) + f(2)\} + 9 \cdot f(3) = (-2)^2 \cdot f(-2) + (-1)^2 \cdot f(-1) + 0^2 \cdot f(0) + 1^2 \cdot f(1) + 2^2 \cdot f(2) + 3^2 \cdot f(3) = \sum_x g(x) \cdot f(x)$.

3 Replace $\int$ by $\sum$ in the proof of Theorem 3.

5 **(a)** $E(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dy\, dx$; **(b)** $E(x) = \int_{-\infty}^{\infty} x g(x)\, dx$.

7 $E(Y) = \frac{37}{12}$.

9 **(a)** 2.4 and 6.24; **(b)** 88.96.

11 $-\frac{11}{6}$.

13 $\frac{1}{2}$.

15 $\frac{1}{12}$.

19 $\mu = \frac{4}{3}, \mu_2' = 2$, and $\sigma^2 = \frac{2}{9}$.

25 $\mu_3 = \mu_3' - \mu\mu_2' + 2\mu^3$ and $\mu_4 = \mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4$.

27 **(a)** 3.2; **(b)** 2.6.

31 **(a)** $k = \sqrt{20}$; **(b)** $k = 10$.

33 $M_X(t) = \frac{2e^t}{3-e^t}, \quad \mu_1' = \frac{3}{2}, \quad \mu_2' = 3, \quad \sigma^2 = \frac{3}{4}$.

35 $\mu = 4, \quad \sigma^2 = 4$.

43 -0.14.

45 $\frac{1}{72}$.

49 **(a)** $\mu_Y = -7, \quad \sigma_Y^2 = 155$; **(b)** $\mu_Z = 19, \quad \sigma_Z^2 = 36$.

51 $\frac{805}{162}$.

53 $\text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y), \quad \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y), \ \text{var}(X) - \text{var}(Y)$.

55 -56.

57 3.

59 $\frac{5}{12}$.

61 **(a)** 98; **(b)** 29,997.

63 \$5.

65 \$3,000.

67 6 million liters.

69 $\frac{a}{a+b}$.

71 $\mu = 4, \quad \sigma^2 = 16$.

73 $\mu = 1, \quad \sigma^2 = 1$.

75 At least $\frac{63}{64}$.

77 0.9179.

79 $\mu = 3.6, \sigma = 0.0224$.

81 **(a)** 0.74, 0.68; **(b)** 1.91, 1.05.

83 0.8.

85 2.95 min.

Special Probability Distributions

1 Introduction

In this chapter we shall study some of the probability distributions that figure most prominently in statistical theory and applications. We shall also study their **parameters**, that is, the quantities that are constants for particular distributions but that can take on different values for different members of families of distributions of the same kind. The most common parameters are the lower moments, mainly μ and σ^2, and there are essentially two ways in which they can be obtained: We can evaluate the necessary sums directly or we can work with moment-generating functions. Although it would seem logical to use in each case whichever method is simplest, we shall sometimes use both. In some instances this will be done because the results are needed later; in others it will merely serve to provide the reader with experience in the application of the respective mathematical techniques. Also, to keep the size of this chapter within bounds, many of the details are left as exercises.

2 The Discrete Uniform Distribution

If a random variable can take on k different values with equal probability, we say that it has a **discrete uniform distribution**; symbolically, we have the following definition.

DEFINITION 1. DISCRETE UNIFORM DISTRIBUTION. *A random variable* X *has a* ***discrete uniform distribution*** *and it is referred to as a discrete uniform random variable if and only if its probability distribution is given by*

$$f(x) = \frac{1}{k} \quad \text{for } x = x_1, x_2, \ldots x_k$$

where $x_i \neq x_j$ *when* $i \neq j$.

From Chapter 5 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

In the special case where $x_i = i$, the discrete uniform distribution becomes $f(x) = \frac{1}{k}$ for $x = 1, 2, \ldots, k$, and in this form it applies, for example, to the number of points we roll with a balanced die. The mean and the variance of this discrete uniform distribution and its moment-generating function are treated in Exercises 1 and 2.

3 The Bernoulli Distribution

If an experiment has two possible outcomes, "success" and "failure," and their probabilities are, respectively, θ and $1-\theta$, then the number of successes, 0 or 1, has a **Bernoulli distribution**; symbolically, we have the following definition.

DEFINITION 2. BERNOULLI DISTRIBUTION. *A random variable* X *has a* ***Bernoulli distribution*** *and it is referred to as a Bernoulli random variable if and only if its probability distribution is given by*

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} \quad \text{for } x = 0, 1$$

Thus, $f(0;\theta) = 1-\theta$ and $f(1;\theta) = \theta$ are combined into a single formula. Observe that we used the notation $f(x;\theta)$ to indicate explicitly that the Bernoulli distribution has the one parameter θ.

In connection with the Bernoulli distribution, a success may be getting heads with a balanced coin, it may be catching pneumonia, it may be passing (or failing) an examination, and it may be losing a race. This inconsistency is a carryover from the days when probability theory was applied only to games of chance (and one player's failure was the other's success). Also for this reason, we refer to an experiment to which the Bernoulli distribution applies as a **Bernoulli trial**, or simply a **trial**, and to sequences of such experiments as **repeated trials**.

4 The Binomial Distribution

Repeated trials play a very important role in probability and statistics, especially when the number of trials is fixed, the parameter θ (the probability of a success) is the same for each trial, and the trials are all independent. As we shall see, several random variables arise in connection with repeated trials. The one we shall study here concerns the total number of successes; others will be given in Section 5.

The theory that we shall discuss in this section has many applications; for instance, it applies if we want to know the probability of getting 5 heads in 12 flips of a coin, the probability that 7 of 10 persons will recover from a tropical disease, or the probability that 35 of 80 persons will respond to a mail-order solicitation. However, this is the case only if each of the 10 persons has the same chance of recovering from the disease and their recoveries are independent (say, they are treated by different doctors in different hospitals), and if the probability of getting a reply to the mail-order solicitation is the same for each of the 80 persons and there is independence (say, no two of them belong to the same household).

To derive a formula for the probability of getting "x successes in n trials" under the stated conditions, observe that the probability of getting x successes and $n-x$ failures *in a specific order* is $\theta^x(1-\theta)^{n-x}$. There is one factor θ for each success, one factor $1-\theta$ for each failure, and the x factors θ and $n-x$ factors $1-\theta$ are all multiplied together by virtue of the assumption of independence. Since this probability applies to any sequence of n trials in which there are x successes and $n-x$ failures, we have only to count how many sequences of this kind there are and then multiply $\theta^x(1-\theta)^{n-x}$ by that number. Clearly, the number of ways in which we can select the x trials on which there is to be a success is $\binom{n}{x}$, and it follows that the desired probability for "x successes in n trials" is $\binom{n}{x}\theta^x(1-\theta)^{n-x}$.

DEFINITION 3. BINOMIAL DISTRIBUTION. *A random variable* X *has a* ***binomial distribution*** *and it is referred to as a binomial random variable if and only if its probability distribution is given by*

$$b(x;n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad \text{for } x = 0,1,2,\ldots n$$

Thus, the number of successes in n trials is a random variable having a binomial distribution with the parameters n and θ. The name "binomial distribution" derives from the fact that the values of $b(x;n,\theta)$ for $x = 0,1,2,\ldots,n$ are the successive terms of the binomial expansion of $[(1-\theta)+\theta]^n$; this shows also that the sum of the probabilities equals 1, as it should.

EXAMPLE 1

Find the probability of getting five heads and seven tails in 12 flips of a balanced coin.

Solution

Substituting $x = 5$, $n = 12$, and $\theta = \frac{1}{2}$ into the formula for the binomial distribution, we get

$$b\left(5;12,\frac{1}{2}\right) = \binom{12}{5}\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^{12-5}$$

and, looking up the value of $\binom{12}{5}$ in Table VII of "Statistical Tables", we find that the result is $792\left(\frac{1}{2}\right)^{12}$, or approximately 0.19.

EXAMPLE 2

Find the probability that 7 of 10 persons will recover from a tropical disease if we can assume independence and the probability is 0.80 that any one of them will recover from the disease.

Solution

Substituting $x = 7$, $n = 10$, and $\theta = 0.80$ into the formula for the binomial distribution, we get

$$b(7; 10, 0.80) = \binom{10}{7}(0.80)^7(1-0.80)^{10-7}$$

and, looking up the value of $\binom{10}{7}$ in Table VII of "Statistical Tables", we find that the result is $120(0.80)^7(0.20)^3$, or approximately 0.20.

If we tried to calculate the third probability asked for on the previous page, the one concerning the responses to the mail-order solicitation, by substituting $x = 35$, $n = 80$, and, say, $\theta = 0.15$, into the formula for the binomial distribution, we would find that this requires a prohibitive amount of work. In actual practice, binomial probabilities are rarely calculated directly, for they are tabulated extensively for various values of θ and n, and there exists an abundance of computer software yielding binomial probabilities as well as the corresponding cumulative probabilities

$$B(x; n, \theta) = \sum_{k=0}^{x} b(k; n, \theta)$$

upon simple commands. An example of such a printout (with somewhat different notation) is shown in Figure 1.

In the past, the National Bureau of Standards table and the book by H. G. Romig have been widely used; they are listed among the references at the end of this chapter. Also, Table I of "Statistical Tables" gives the values of $b(x; n, \theta)$ to four decimal places for $n = 1$ to $n = 20$ and $\theta = 0.05, 0.10, 0.15, \ldots, 0.45, 0.50$. To use this table when θ is greater than 0.50, we refer to the following identity.

```
MTB > BINOMIAL N=1Ø  P=Ø.63

   BINOMIAL PROBABILITIES FOR N = 1Ø AND P = .63ØØØØ

      K               P(X = K)              P(X LESS OR = K)
      Ø                 .ØØØØ                   .ØØØØ
      1                 .ØØØ8                   .ØØØ9
      2                 .ØØ63                   .ØØ71
      3                 .Ø285                   .Ø356
      4                 .Ø849                   .12Ø5
      5                 .1734                   .2939
      6                 .2461                   .54ØØ
      7                 .2394                   .7794
      8                 .1529                   .9323
      9                 .Ø578                   .99Ø2
     10                 .ØØ98                  1.ØØØØ
```

Figure 1. Computer printout of binomial probabilities for $n = 10$ and $\theta = 0.63$.

THEOREM 1.

$$b(x; n, \theta) = b(n - x; n, 1 - \theta)$$

which the reader will be asked to prove in part (a) of Exercise 5. For instance, to find $b(11; 18, 0.70)$, we look up $b(7; 18, 0.30)$ and get 0.1376. Also, there are several ways in which binomial probabilities can be approximated when n is large; one of these will be mentioned in Section 7.

Let us now find formulas for the mean and the variance of the binomial distribution.

THEOREM 2. The mean and the variance of the binomial distribution are

$$\mu = n\theta \quad \text{and} \quad \sigma^2 = n\theta(1 - \theta)$$

Proof

$$\begin{aligned}\mu &= \sum_{x=0}^{n} x \cdot \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} \theta^x (1-\theta)^{n-x}\end{aligned}$$

where we omitted the term corresponding to $x = 0$, which is 0, and canceled the x against the first factor of $x! = x(x-1)!$ in the denominator of $\binom{n}{x}$. Then, factoring out the factor n in $n! = n(n-1)!$ and one factor θ, we get

$$\mu = n\theta \cdot \sum_{x=1}^{n} \binom{n-1}{x-1} \theta^{x-1} (1-\theta)^{n-x}$$

and, letting $y = x - 1$ and $m = n - 1$, this becomes

$$\mu = n\theta \cdot \sum_{y=0}^{m} \binom{m}{y} \theta^y (1-\theta)^{m-y} = n\theta$$

since the last summation is the sum of all the values of a binomial distribution with the parameters m and θ, and hence equal to 1.

To find expressions for μ_2' and σ^2, let us make use of the fact that $E(X^2) = E[X(X-1)] + E(X)$ and first evaluate $E[X(X-1)]$. Duplicating for all practical purposes the steps used before, we thus get

$$\begin{aligned}E[X(X-1)] &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} \theta^x (1-\theta)^{n-x} \\ &= n(n-1)\theta^2 \cdot \sum_{x=2}^{n} \binom{n-2}{x-2} \theta^{x-2} (1-\theta)^{n-x}\end{aligned}$$

and, letting $y = x-2$ and $m = n-2$, this becomes

$$E[X(X-1)] = n(n-1)\theta^2 \cdot \sum_{y=0}^{m} \binom{m}{y} \theta^y (1-\theta)^{m-y}$$
$$= n(n-1)\theta^2$$

Therefore,

$$\mu_2' = E[X(X-1)] + E(X) = n(n-1)\theta^2 + n\theta$$

and, finally,

$$\begin{aligned}\sigma^2 &= \mu_2' - \mu^2 \\ &= n(n-1)\theta^2 + n\theta - n^2\theta^2 \\ &= n\theta(1-\theta)\end{aligned}$$

An alternative proof of this theorem, requiring much less algebraic detail, is suggested in Exercise 6.

It should not have come as a surprise that the mean of the binomial distribution is given by the product $n\theta$. After all, if a balanced coin is flipped 200 times, we expect (in the sense of a mathematical expectation) $200 \cdot \frac{1}{2} = 100$ heads and 100 tails; similarly, if a balanced die is rolled 240 times, we expect $240 \cdot \frac{1}{6} = 40$ sixes, and if the probability is 0.80 that a person shopping at a department store will make a purchase, we would expect $400(0.80) = 320$ of 400 persons shopping at the department store to make a purchase.

The formula for the variance of the binomial distribution, being a measure of variation, has many important applications; but, to emphasize its significance, let us consider the random variable $Y = \frac{X}{n}$, where X is a random variable having a binomial distribution with the parameters n and θ. This random variable is the proportion of successes in n trials, and in Exercise 6 the reader will be asked to prove the following result.

THEOREM 3. If X has a binomial distribution with the parameters n and θ and $Y = \frac{X}{n}$, then

$$E(Y) = \theta \quad \text{and} \quad \sigma_Y^2 = \frac{\theta(1-\theta)}{n}$$

Now, if we apply Chebyshev's theorem with $k\sigma = c$, we can assert that *for any positive constant* c *the probability is at least*

$$1 - \frac{\theta(1-\theta)}{nc^2}$$

that the proportion of successes in n *trials falls between* $\theta - c$ *and* $\theta + c$. Hence, *when* $n \to \infty$, *the probability approaches* 1 *that the proportion of successes will differ from* θ *by less than any arbitrary constant* c. This result is called a **law of large numbers**, and it should be observed that it applies to the proportion of successes, not to their actual number. It is a fallacy to suppose that when n is large the number of successes must necessarily be close to $n\theta$.

```
MTB > BRANDOM 1ØØ N=1  P=.5 C1

   1ØØ BINOMIAL EXPERIMENTS WITH N = 1 AND P = .5ØØØ
     Ø.    Ø.     1.   1.   1.    1.   1.   Ø.   Ø.    1.
     1.    Ø.     Ø.   1.   Ø.    1.   1.   1.   Ø.    1.
     Ø.    Ø.     1.   Ø.   1.    1.   Ø.   1.   Ø.    Ø.
     1.    1.     Ø.   1.   Ø.    Ø.   1.   1.   1.    Ø.
     1.    Ø.     1.   Ø.   Ø.    Ø.   Ø.   1.   Ø.    Ø.
     1.    1.     Ø.   Ø.   Ø.    Ø.   Ø.   1.   Ø.    Ø.
     1.    1.     Ø.   Ø.   1.    1.   1.   Ø.   1.    1.
     1.    Ø.     1.   1.   Ø.    1.   1.   Ø.   Ø.    Ø.
     Ø.    Ø.     Ø.   1.   Ø.    Ø.   1.   Ø.   1.    1.
     1.    Ø.     1.   1.   1.    1.   Ø.   1.   Ø.    1.
   SUMMARY

   VALUE  FREQUENCY
     Ø      49
     1      51
```

Figure 2. Computer simulation of 100 flips of a balanced coin.

An easy illustration of this law of large numbers can be obtained through a **computer simulation** of the repeated flipping of a balanced coin. This is shown in Figure 2, where the 1's and 0's denote heads and tails.

Reading across successive rows, we find that among the first five simulated flips there are 3 heads, among the first ten there are 6 heads, among the first fifteen there are 8 heads, among the first twenty there are 12 heads, among the first twenty-five there are 14 heads, . . ., and among all hundred there are 51 heads. The corresponding proportions, plotted in Figure 3, are $\frac{3}{5} = 0.60$, $\frac{6}{10} = 0.60$, $\frac{8}{15} = 0.53$, $\frac{12}{20} = 0.60$, $\frac{14}{25} = 0.56, \ldots$, and $\frac{51}{100} = 0.51$. Observe that the proportion of heads fluctuates but comes closer and closer to 0.50, the probability of heads for each flip of the coin.

Since the moment-generating function of the binomial distribution is easy to obtain, let us find it and use it to verify the results of Theorem 2.

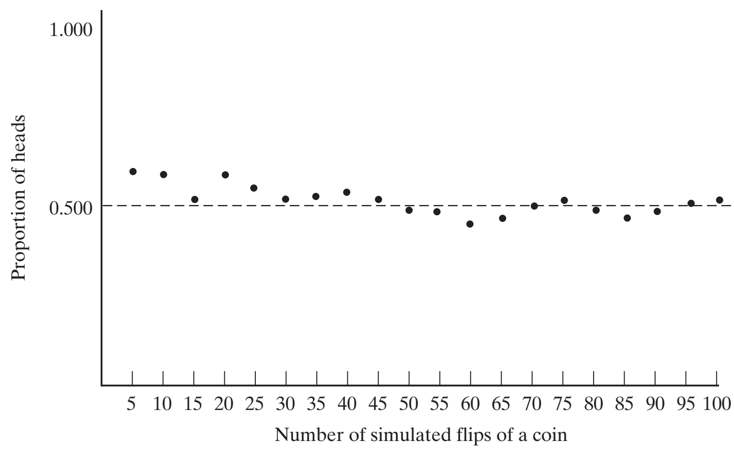

Figure 3. Graph illustrating the law of large numbers.

THEOREM 4. The moment-generating function of the binomial distribution is given by

$$M_X(t) = [1+\theta(e^t-1)]^n$$

If we differentiate $M_X(t)$ twice with respect to t, we get

$$\begin{aligned} M'_X(t) &= n\theta e^t[1+\theta(e^t-1)]^{n-1} \\ M''_X(t) &= n\theta e^t[1+\theta(e^t-1)]^{n-1} + n(n-1)\theta^2 e^{2t}[1+\theta(e^t-1)]^{n-2} \\ &= n\theta e^t(1-\theta+n\theta e^t)[1+\theta(e^t-1)]^{n-2} \end{aligned}$$

and, upon substituting $t=0$, we get $\mu'_1 = n\theta$ and $\mu'_2 = n\theta(1-\theta+n\theta)$. Thus, $\mu = n\theta$ and $\sigma^2 = \mu'_2 - \mu^2 = n\theta(1-\theta+n\theta)-(n\theta)^2 = n\theta(1-\theta)$, which agrees with the formulas given in Theorem 2.

From the work of this section it may seem easier to find the moments of the binomial distribution with the moment-generating function than to evaluate them directly, but it should be apparent that the differentiation becomes fairly involved if we want to determine, say, μ'_3 or μ'_4. Actually, there exists yet an easier way of determining the moments of the binomial distribution; it is based on its **factorial moment-generating function**, which is explained in Exercise 12.

Exercises

1. If X has the discrete uniform distribution $f(x) = \frac{1}{k}$ for $x = 1, 2, \ldots, k$, show that
(a) its mean is $\mu = \frac{k+1}{2}$;
(b) its variance is $\sigma^2 = \frac{k^2-1}{12}$.

2. If X has the discrete uniform distribution $f(x) = \frac{1}{k}$ for $x = 1, 2, \ldots, k$, show that its moment-generating function is given by

$$M_X(t) = \frac{e^t(1-e^{kt})}{k(1-e^t)}$$

Also find the mean of this distribution by evaluating $\lim_{t\to 0} M'_X(t)$, and compare the result with that obtained in Exercise 1.

3. We did not study the Bernoulli distribution in any detail in Section 3, because it can be looked upon as a binomial distribution with $n = 1$. Show that for the Bernoulli distribution, $\mu'_r = \theta$ for $r = 1, 2, 3, \ldots$, by
(a) evaluating the sum $\sum_{x=0}^{1} x^r \cdot f(x;\theta)$;
(b) letting $n = 1$ in the moment-generating function of the binomial distribution and examining its Maclaurin's series.

4. This question has been intentionally omitted for this edition.

5. Verify that
(a) $b(x;n,\theta) = b(n-x;n,1-\theta)$.
Also show that if $B(x;n,\theta) = \sum_{k=0}^{x} b(k;n,\theta)$ for $x = 0, 1, 2, \ldots, n$, then
(b) $b(x;n,\theta) = B(x;n,\theta) - B(x-1;n,\theta)$;
(c) $b(x;n,\theta) = B(n-x;n,1-\theta) - B(n-x-1;n,1-\theta)$;
(d) $B(x;n,\theta) = 1 - B(n-x-1;n,1-\theta)$.

6. An alternative proof of Theorem 2 may be based on the fact that if $X_1, X_2, \ldots$, and X_n are independent random variables having the same Bernoulli distribution with the parameter θ, then $Y = X_1 + X_2 + \cdots + X_n$ is a random variable having the binomial distribution with the parameters n and θ.

Verify directly (that is, without making use of the fact that the Bernoulli distribution is a special case of the binomial distribution) that the mean and the variance of the Bernoulli distribution are $\mu = \theta$ and $\sigma^2 = \theta(1-\theta)$.

7. Prove Theorem 3.

8. When calculating all the values of a binomial distribution, the work can usually be simplified by first calculating $b(0;n,\theta)$ and then using the recursion formula

$$b(x+1;n,\theta) = \frac{\theta(n-x)}{(x+1)(1-\theta)} \cdot b(x;n,\theta)$$

Verify this formula and use it to calculate the values of the binomial distribution with $n = 7$ and $\theta = 0.25$.

9. Use the recursion formula of Exercise 8 to show that for $\theta = \frac{1}{2}$ the binomial distribution has
(a) a maximum at $x = \dfrac{n}{2}$ when n is even;

(b) maxima at $x = \dfrac{n-1}{2}$ and $x = \dfrac{n+1}{2}$ when n is odd.

10. If X is a binomial random variable, for what value of θ is the probability $b(x; n, \theta)$ a maximum?

11. In the proof of Theorem 2 we determined the quantity $E[X(X-1)]$, called the second **factorial moment**. In general, the rth factorial moment of X is given by

$$\mu'_{(r)} = E[X(X-1)(X-2)\cdot \ldots \cdot (X-r+1)]$$

Express μ'_2, μ'_3, and μ'_4 in terms of factorial moments.

12. The **factorial moment-generating function** of a discrete random variable X is given by

$$F_X(t) = E(t^X) = \sum_x t^x \cdot f(x)$$

Show that the rth derivative of $F_X(t)$ with respect to t at $t = 1$ is $\mu'_{(r)}$, the rth factorial moment defined in Exercise 11.

13. With reference to Exercise 12, find the factorial moment-generating function of
(a) the Bernoulli distribution and show that $\mu'_{(1)} = \theta$ and $\mu'_{(r)} = 0$ for $r > 1$;
(b) the binomial distribution and use it to find μ and σ^2.

14. This question has been intentionally omitted for this edition.

15. This question has been intentionally omitted for this edition.

5 The Negative Binomial and Geometric Distributions

In connection with repeated Bernoulli trials, we are sometimes interested in the number of the trial on which the kth success occurs. For instance, we may be interested in the probability that the tenth child exposed to a contagious disease will be the third to catch it, the probability that the fifth person to hear a rumor will be the first one to believe it, or the probability that a burglar will be caught for the second time on his or her eighth job.

If the kth success is to occur on the xth trial, there must be $k-1$ successes on the first $x-1$ trials, and the probability for this is

$$b(k-1; x-1, \theta) = \binom{x-1}{k-1} \theta^{k-1}(1-\theta)^{x-k}$$

The probability of a success on the xth trial is θ, and the probability that the kth success occurs on the xth trial is, therefore,

$$\theta \cdot b(k-1; x-1, \theta) = \binom{x-1}{k-1} \theta^{k}(1-\theta)^{x-k}$$

Definition 4. negative binomial distribution. *A random variable* X *has a* ***negative binomial distribution*** *and it is referred to as a negative binomial random variable if and only if*

$$b^*(x; k, \theta) = \binom{x-1}{k-1} \theta^{k}(1-\theta)^{x-k} \quad \text{for } x = k, k+1, k+2, \ldots.$$

Thus, the number of the trial on which the kth success occurs is a random variable having a negative binomial distribution with the parameters k and θ. The name "negative binomial distribution" derives from the fact that the values of $b^*(x; k, \theta)$ for $x = k, k+1, k+2, \ldots$ are the successive terms of the binomial expansion of

$\left(\frac{1}{\theta} - \frac{1-\theta}{\theta}\right)^{-k}$.[†] In the literature of statistics, negative binomial distributions are also referred to as **binomial waiting-time distributions** or as **Pascal distributions**.

EXAMPLE 3

If the probability is 0.40 that a child exposed to a certain contagious disease will catch it, what is the probability that the tenth child exposed to the disease will be the third to catch it?

Solution

Substituting $x = 10$, $k = 3$, and $\theta = 0.40$ into the formula for the negative binomial distribution, we get

$$\begin{aligned} b^*(10; 3, 0.40) &= \binom{9}{2}(0.40)^3(0.60)^7 \\ &= 0.0645 \end{aligned}$$

When a table of binomial probabilities is available, the determination of negative binomial probabilities can generally be simplified by making use of the following identity.

> **THEOREM 5.**
> $$b^*(x; k, \theta) = \frac{k}{x} \cdot b(k; x, \theta)$$

The reader will be asked to verify this theorem in Exercise 18.

EXAMPLE 4

Use Theorem 5 and Table I of "Statistical Tables" to rework Example 3.

Solution

Substituting $x = 10$, $k = 3$, and $\theta = 0.40$ into the formula of Theorem 5, we get

$$\begin{aligned} b^*(10; 3, 0.40) &= \frac{3}{10} \cdot b(3; 10, 0.40) \\ &= \frac{3}{10}(0.2150) \\ &= 0.0645 \end{aligned}$$

Moments of the negative binomial distribution may be obtained by proceeding as in the proof of Theorem 2; for the mean and the variance we obtain the following theorem.

> **THEOREM 6.** The mean and the variance of the negative binomial distribution are
> $$\mu = \frac{k}{\theta} \quad \text{and} \quad \sigma^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right)$$

as the reader will be asked to verify in Exercise 19.

[†] Binomial expansions with negative exponents are explained in Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. New York: John Wiley & Sons, Inc., 1968.

Since the negative binomial distribution with $k = 1$ has many important applications, it is given a special name; it is called the **geometric distribution**.

DEFINITION 5. GEOMETRIC DISTRIBUTION. *A random variable* X *has a* ***geometric distribution*** *and it is referred to as a geometric random variable if and only if its probability distribution is given by*

$$g(x; \theta) = \theta(1-\theta)^{x-1} \quad \text{for } x = 1, 2, 3, \ldots$$

EXAMPLE 5

If the probability is 0.75 that an applicant for a driver's license will pass the road test on any given try, what is the probability that an applicant will finally pass the test on the fourth try?

Solution

Substituting $x = 4$ and $\theta = 0.75$ into the formula for the geometric distribution, we get

$$\begin{aligned} g(4; 0.75) &= 0.75(1-0.75)^{4-1} \\ &= 0.75(0.25)^3 \\ &= 0.0117 \end{aligned}$$

Of course, this result is based on the assumption that the trials are all independent, and there may be some question here about its validity.

6 The Hypergeometric Distribution

To obtain a formula analogous to that of the binomial distribution that applies to sampling without replacement, in which case the trials are not independent, let us consider a set of N elements of which M are looked upon as successes and the other $N-M$ as failures. As in connection with the binomial distribution, we are interested in the probability of getting x successes in n trials, but now we are choosing, without replacement, n of the N elements contained in the set.

There are $\binom{M}{x}$ ways of choosing x of the M successes and $\binom{N-M}{n-x}$ ways of choosing $n-x$ of the $N-M$ failures, and, hence, $\binom{M}{x}\binom{N-M}{n-x}$ ways of choosing x successes and $n-x$ failures. Since there are $\binom{N}{n}$ ways of choosing n of the N elements in the set, and we shall assume that they are all equally likely (which is what we mean when we say that the selection is random), the probability of "x successes in n trials" is $\binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n}$.

DEFINITION 6. HYPERGEOMETRIC DISTRIBUTION. *A random variable* X *has a* ***hypergeometric distribution*** *and it is referred to as a hypergeometric random variable if and only if its probability distribution is given by*

$$h(x; n, N, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \text{for } x = 0, 1, 2, \ldots, n \quad x \le M \text{ and } n - x \le N - M$$

Thus, for sampling without replacement, the number of successes in n trials is a random variable having a hypergeometric distribution with the parameters n, N, and M.

EXAMPLE 6

As part of an air-pollution survey, an inspector decides to examine the exhaust of 6 of a company's 24 trucks. If 4 of the company's trucks emit excessive amounts of pollutants, what is the probability that none of them will be included in the inspector's sample?

Solution

Substituting $x = 0$, $n = 6$, $N = 24$, and $M = 4$ into the formula for the hypergeometric distribution, we get

$$h(0; 6, 24, 4) = \frac{\binom{4}{0}\binom{20}{6}}{\binom{24}{6}}$$

$$= 0.2880$$

The method by which we find the mean and the variance of the hypergeometric distribution is very similar to that employed in the proof of Theorem 2.

THEOREM 7. The mean and the variance of the hypergeometric distribution are

$$\mu = \frac{nM}{N} \quad \text{and} \quad \sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

Proof To determine the mean, let us directly evaluate the sum

$$\mu = \sum_{x=0}^{n} x \cdot \frac{\binom{M}{x}\left(N - Mn - x\right)}{\binom{N}{n}}$$

$$= \sum_{x=1}^{n} \frac{M!}{(x-1)!(M-x)!} \cdot \frac{\binom{N-M}{n-x}}{\binom{N}{n}}$$

where we omitted the term corresponding to $x = 0$, which is 0, and canceled the x against the first factor of $x! = x(x-1)!$ in the denominator of $\binom{M}{x}$. Then, factoring out $M \Big/ \binom{N}{n}$, we get

$$\mu = \frac{M}{\binom{N}{n}} \cdot \sum_{x=1}^{n} \binom{M-1}{x-1} \binom{N-M}{n-x}$$

and, letting $y = x - 1$ and $m = n - 1$, this becomes

$$\mu = \frac{M}{\binom{N}{n}} \cdot \sum_{y=0}^{m} \binom{M-1}{y} \binom{N-M}{m-y}$$

Finally, using $\sum_{r=0}^{k} \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}$, we get

$$\mu = \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{m} = \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{n-1} = \frac{nM}{N}$$

To obtain the formula for σ^2, we proceed as in the proof of Theorem 2 by first evaluating $E[X(X-1)]$ and then making use of the fact that $E(X^2) = E[X(X-1)] + E(X)$. Leaving it to the reader to show that

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

in Exercise 27, we thus get

$$\begin{aligned}\sigma^2 &= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2 \\ &= \frac{nM(N-M)(N-n)}{N^2(N-1)}\end{aligned}$$

The moment-generating function of the hypergeometric distribution is fairly complicated. Details of this may be found in the book *The Advanced Theory of Statistics* by M. G. Kendall and A. Stuart.

When N is large and n is relatively small compared to N (the usual rule of thumb is that n should not exceed 5 percent of N), there is not much difference between sampling with replacement and sampling without replacement, and the formula for the binomial distribution with the parameters n and $\theta = \frac{M}{N}$ may be used to approximate hypergeometric probabilities.

EXAMPLE 7

Among the 120 applicants for a job, only 80 are actually qualified. If 5 of the applicants are randomly selected for an in-depth interview, find the probability that only 2 of the 5 will be qualified for the job by using

(a) the formula for the hypergeometric distribution;

(b) the formula for the binomial distribution with $\theta = \frac{80}{120}$ as an approximation.

Solution

(a) Substituting $x = 2$, $n = 5$, $N = 120$, and $M = 80$ into the formula for the hypergeometric distribution, we get

$$h(2; 5, 120, 80) = \frac{\binom{80}{2}\binom{40}{3}}{\binom{120}{5}}$$
$$= 0.164$$

rounded to three decimals;

(b) substituting $x = 2$, $n = 5$, and $\theta = \frac{80}{120} = \frac{2}{3}$ into the formula for the binomial distribution, we get

$$b\left(2; 5, \frac{2}{3}\right) = \binom{5}{2}\left(\frac{2}{3}\right)^2\left(1 - \frac{2}{3}\right)^3$$
$$= 0.165$$

rounded to three decimals. As can be seen from these results, the approximation is very close.

7 The Poisson Distribution

When n is large, the calculation of binomial probabilities with the formula of Definition 3 will usually involve a prohibitive amount of work. For instance, to calculate the probability that 18 of 3,000 persons watching a parade on a very hot summer day will suffer from heat exhaustion, we first have to determine $\binom{3{,}000}{18}$, and if the probability is 0.005 that any one of the 3,000 persons watching the parade will suffer from heat exhaustion, we also have to calculate the value of $(0.005)^{18}(0.995)^{2{,}982}$.

In this section we shall present a probability distribution that can be used to approximate binomial probabilities of this kind. Specifically, we shall investigate the limiting form of the binomial distribution when $n \to \infty$, $\theta \to 0$, while $n\theta$ remains constant. Letting this constant be λ, that is, $n\theta = \lambda$ and, hence, $\theta = \frac{\lambda}{n}$, we can write

$$b(x;n,\theta) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$
$$= \frac{n(n-1)(n-2)\cdot\ldots\cdot(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Then, if we divide one of the x factors n in $\left(\frac{\lambda}{n}\right)^x$ into each factor of the product $n(n-1)(n-2)\cdot\ldots\cdot(n-x+1)$ and write

$$\left(1-\frac{\lambda}{n}\right)^{n-x} \quad \text{as} \quad \left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-x}$$

we obtain

$$\frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{x-1}{n}\right)}{x!}(\lambda)^x\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-x}$$

Finally, if we let $n \to \infty$ while x and λ remain fixed, we find that

$$1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{x-1}{n}\right) \to 1$$
$$\left(1-\frac{\lambda}{n}\right)^{-x} \to 1$$
$$\left(1-\frac{\lambda}{n}\right)^{-n/\lambda} \to e$$

and, hence, that the limiting distribution becomes

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

DEFINITION 7. POISSON DISTRIBUTION. *A random variable has a **Poisson distribution** and it is referred to as a Poisson random variable if and only if its probability distribution is given by*

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

Thus, in the limit when $n \to \infty, \theta \to 0$, and $n\theta = \lambda$ remains constant, the number of successes is a random variable having a Poisson distribution with the parameter λ. This distribution is named after the French mathematician Simeon Poisson (1781–1840). In general, the Poisson distribution will provide a good approximation to binomial probabilities when $n \geqq 20$ and $\theta \leqq 0.05$. When $n \geqq 100$ and $n\theta < 10$, the approximation will generally be excellent.

To get some idea about the closeness of the Poisson approximation to the binomial distribution, consider the computer printout of Figure 4, which shows, one above the other, the binomial distribution with $n = 150$ and $\theta = 0.05$ and the Poisson distribution with $\lambda = 150(0.05) = 7.5$.

EXAMPLE 8

Use Figure 4 to determine the value of x (from 5 to 15) for which the error is greatest when we use the Poisson distribution with $\lambda = 7.5$ to approximate the binomial distribution with $n = 150$ and $\theta = 0.05$.

Solution
Calculating the differences corresponding to $x = 5, x = 6, \ldots, x = 15$, we get 0.0006, −0.0017, −0.0034, −0.0037, −0.0027, −0.0011, 0.0003, 0.0011, 0.0013, 0.0011, and 0.0008. Thus, the maximum error (numerically) is −0.0037, and it corresponds to $x = 8$.

The examples that follow illustrate the Poisson approximation to the binomial distribution.

EXAMPLE 9

If 2 percent of the books bound at a certain bindery have defective bindings, use the Poisson approximation to the binomial distribution to determine the probability that 5 of 400 books bound by this bindery will have defective bindings.

Solution
Substituting $x = 5$, $\lambda = 400(0.02) = 8$, and $e^{-8} = 0.00034$ (from Table VIII of "Statistical Tables") into the formula of Definition 7, we get

$$p(5; 8) = \frac{8^5 \cdot e^{-8}}{5!} = \frac{(32{,}768)(0.00034)}{120} = 0.093$$

In actual practice, Poisson probabilities are seldom obtained by direct substitution into the formula of Definition 7. Sometimes we refer to tables of Poisson probabilities, such as Table II of "Statistical Tables", or more extensive tables in handbooks of statistical tables, but more often than not, nowadays, we refer to suitable computer software. The use of tables or computers is of special importance when we are concerned with probabilities relating to several values of x.

EXAMPLE 10

Records show that the probability is 0.00005 that a car will have a flat tire while crossing a certain bridge. Use the Poisson distribution to approximate the binomial probabilities that, among 10,000 cars crossing this bridge,

(a) exactly two will have a flat tire;
(b) at most two will have a flat tire.

Solution

(a) Referring to Table II of "Statistical Tables", we find that for $x = 2$ and $\lambda = 10,000(0.00005) = 0.5$, the Poisson probability is 0.0758.
(b) Referring to Table II of "Statistical Tables", we find that for $x = 0$, 1, and 2, and $\lambda = 0.5$, the Poisson probabilities are 0.6065, 0.3033, and 0.0758. Thus, the

```
MTB > BINOMIAL N=15Ø  P=Ø.Ø5
  BINOMIAL PROBABILITIES FOR N = 1.5Ø AND P = .Ø5ØØØØ

    K       P(X = K)    P(X LESS OR = K)
    Ø        .ØØØ5          .ØØØ5
    1        .ØØ36          .ØØ41
    2        .Ø141          .Ø182
    3        .Ø366          .Ø548
    4        .Ø7Ø8          .1256
    5        .1Ø88          .2344
    6        .1384          .3729
    7        .1499          .5228
    8        .141Ø          .6638
    9        .1171          .78Ø9
   10        .Ø869          .8678
   11        .Ø582          .926Ø
   12        .Ø355          .9615
   13        .Ø198          .9813
   14        .Ø1Ø2          .9915
   15        .ØØ49          .9964
   16        .ØØ22          .9986
   17        .ØØØ9          .9995
   18        .ØØØ3          .9998
   19        .ØØØ1          .9999


MTB > POISSON MU=7.5
  POISSON PROBABILITIES FOR MEAN = 7.5ØØ

    K       P(X = K)    P(X LESS OR = K)
    Ø        .ØØØ6          .ØØØ6
    1        .ØØ41          .ØØ47
    2        .Ø156          .Ø2Ø3
    3        .Ø389          .Ø591
    4        .Ø729          .1321
    5        .1Ø94          .2414
    6        .1367          .3782
    7        .1465          .5246
    8        .1373          .662Ø
    9        .1144          .7764
   10        .Ø858          .8622
   11        .Ø585          .92Ø8
   12        .Ø366          .9573
   13        .Ø211          .9784
   14        .Ø113          .9897
   15        .ØØ57          .9954
   16        .ØØ26          .998Ø
   17        .ØØ12          .9992
   18        .ØØØ5          .9997
   19        .ØØØ2          .9999
   20        .ØØØ1         1.ØØØØ
```

Figure 4. Computer printout of the binomial distribution with $n = 150$ and $\theta = 0.05$ and the Poisson distribution with $\lambda = 7.5$.

probability that at most 2 of 10,000 cars crossing the bridge will have a flat tire is

$$0.6065 + 0.3033 + 0.0758 = 0.9856$$

EXAMPLE 11

Use Figure 5 to rework the preceding example.

Solution

(a) Reading off the value for $K = 2$ in the P(X = K) column, we get 0.0758.

(b) Here we could add the values for $K = 0$, $K = 1$, and $K = 2$ in the P(X = K) column, or we could read the value for $K = 2$ in the P(X LESS OR = K) column, getting 0.9856.

Having derived the Poisson distribution as a limiting form of the binomial distribution, we can obtain formulas for its mean and its variance by applying the same limiting conditions ($n \to \infty$, $\theta \to 0$, and $n\theta = \lambda$ remains constant) to the mean and the variance of the binomial distribution. For the mean we get $\mu = n\theta = \lambda$ and for the variance we get $\sigma^2 = n\theta(1-\theta) = \lambda(1-\theta)$, which approaches λ when $\theta \to 0$.

THEOREM 8. The mean and the variance of the Poisson distribution are given by

$$\mu = \lambda \quad \text{and} \quad \sigma^2 = \lambda$$

These results can also be obtained by directly evaluating the necessary summations (see Exercise 33) or by working with the moment-generating function given in the following theorem.

THEOREM 9. The moment-generating function of the Poisson distribution is given by

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

Proof By Definition 7 and the definition of moment-generating function—*The **moment generating function** of a random variable* X, *where it exists, is given by* $M_X(t) = E(e^{tX}) = \sum_x e^{tX} \cdot f(x)$ *when* X *is discrete, and*

```
MTB > POISSON MU=.5

   POISSON PROBABILITIES FOR MEAN = .5ØØ

     K     P(X = K)  P(X LESS OR = K)
     Ø     .6Ø65             .6Ø65
     1     .3Ø33             .9Ø98
     2     .Ø758             .9856
     3     .Ø126             .9982
     4     .ØØ16             .9998
     5     .ØØØ2            1.ØØØØ
```

Figure 5. Computer printout of the Poisson distribution with $\lambda = 0.5$.

$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx$ *when* X *is continuous*—we get

$$M_X(t) = \sum_{x=0}^{\infty} e^{xt} \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

where $\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$ can be recognized as the Maclaurin's series of e^z with $z = \lambda e^t$. Thus,

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

Then, if we differentiate $M_X(t)$ twice with respect to t, we get

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

so that $\mu'_1 = M'_X(0) = \lambda$ and $\mu'_2 = M''_X(0) = \lambda + \lambda^2$. Thus, $\mu = \lambda$ and $\sigma^2 = \mu'_2 - \mu^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$, which agrees with Theorem 8.

Although the Poisson distribution has been derived as a limiting form of the binomial distribution, it has many applications that have no direct connection with binomial distributions. For example, the Poisson distribution can serve as a model for the number of successes that occur during a given time interval or in a specified region when (1) the numbers of successes occurring in nonoverlapping time intervals or regions are independent, (2) the probability of a single success occurring in a very short time interval or in a very small region is proportional to the length of the time interval or the size of the region, and (3) the probability of more than one success occurring in such a short time interval or falling in such a small region is negligible. Hence, a Poisson distribution might describe the number of telephone calls per hour received by an office, the number of typing errors per page, or the number of bacteria in a given culture when the average number of successes, λ, for the given time interval or specified region is known.

EXAMPLE 12

The average number of trucks arriving on any one day at a truck depot in a certain city is known to be 12. What is the probability that on a given day fewer than 9 trucks will arrive at this depot?

Solution

Let X be the number of trucks arriving on a given day. Then, using Table II of "Statistical Tables" with $\lambda = 12$, we get

$$P(X < 9) = \sum_{x=0}^{8} p(x; 12) = 0.1550$$

If, in a situation where the preceding conditions apply, successes occur at a mean rate of α per *unit* time or per *unit* region, then the number of successes in an interval of t units of time or t units of the specified region is a Poisson random variable with the mean $\lambda = \alpha t$ (see Exercise 31). Therefore, the number of successes, X, in a time interval of length t units or a region of size t units has the Poisson distribution

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

EXAMPLE 13

A certain kind of sheet metal has, on the average, five defects per 10 square feet. If we assume a Poisson distribution, what is the probability that a 15-square-foot sheet of the metal will have at least six defects?

Solution

Let X denote the number of defects in a 15-square-foot sheet of the metal. Then, since the unit of area is 10 square feet, we have

$$\lambda = \alpha t = (5)(1.5) = 7.5$$

and

$$P(X \geqq 6) = 1 - P(X \leqq 5) = 1 - 0.2414 = 0.7586$$

according to the computer printout shown in Figure 4.

Exercises

16. The negative binomial distribution is sometimes defined in a different way as the distribution of the number of failures that precede the kth success. If the kth success occurs on the xth trial, it must be preceded by $x - k$ failures. Thus, find the distribution of $Y = X - k$, where X has the distribution of Definition 4.

17. With reference to Exercise 16, find expressions for μ_Y and σ_Y^2.

18. Prove Theorem 5.

19. Prove Theorem 6 by first determining $E(X)$ and $E[X(X+1)]$.

20. Show that the moment-generating function of the geometric distribution is given by

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

21. Use the moment-generating function derived in Exercise 20 to show that for the geometric distribution, $\mu = \frac{1}{\theta}$ and $\sigma^2 = \frac{1-\theta}{\theta^2}$.

22. Differentiating with respect to θ the expressions on both sides of the equation

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

show that the mean of the geometric distribution is given by $\mu = \frac{1}{\theta}$. Then, differentiating again with respect to θ, show that $\mu_2' = \frac{2-\theta}{\theta^2}$ and hence that $\sigma^2 = \frac{1-\theta}{\theta^2}$.

23. If X is a random variable having a geometric distribution, show that

$$P(X = x + n | X > n) = P(X = x)$$

24. If the probability is $f(x)$ that a product fails the xth time it is being used, that is, on the xth trial, then its **failure rate** at the xth trial is the probability that it will fail on the xth trial given that it has not failed on the first $x - 1$ trials; symbolically, it is given by

$$Z(x) = \frac{f(x)}{1 - F(x-1)}$$

where $F(x)$ is the value of the corresponding distribution function at x. Show that if X is a geometric random variable, its failure rate is constant and equal to θ.

25. A variation of the binomial distribution arises when the n trials are all independent, but the probability of a success on the ith trial is θ_i, and these probabilities are not all equal. If X is the number of successes obtained under these conditions in n trials, show that

(a) $\mu_X = n\theta$, where $\theta = \frac{1}{n} \cdot \sum_{i=1}^{n} \theta_i$;

(b) $\sigma_X^2 = n\theta(1-\theta) - n\sigma_\theta^2$, where θ is as defined in part (a) and $\sigma_\theta^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\theta_i - \theta)^2$.

26. When calculating all the values of a hypergeometric distribution, the work can often be simplified by first calculating $h(0; n, N, M)$ and then using the recursion formula

$$h(x+1; n, N, M) = \frac{(n-x)(M-x)}{(x+1)(N-M-n+x+1)} \cdot h(x; n, N, M)$$

Verify this formula and use it to calculate the values of the hypergeometric distribution with $n = 4$, $N = 9$, and $M = 5$.

27. Verify the expression given for $E[X(X-1)]$ in the proof of Theorem 7.

28. Show that if we let $\theta = \frac{M}{N}$ in Theorem 7, the mean and the variance of the hypergeometric distribution can be written as $\mu = n\theta$ and $\sigma^2 = n\theta(1-\theta) \cdot \frac{N-n}{N-1}$. How do these results tie in with the discussion in the theorem?

29. When calculating all the values of a Poisson distribution, the work can often be simplified by first calculating $p(0; \lambda)$ and then using the recursion formula

$$p(x+1; \lambda) = \frac{\lambda}{x+1} \cdot p(x; \lambda)$$

Verify this formula and use it and $e^{-2} = 0.1353$ to verify the values given in Table II of "Statistical Tables" for $\lambda = 2$.

30. Approximate the binomial probability $b(3; 100, 0.10)$ by using

(a) the formula for the binomial distribution and logarithms;

(b) Table II of "Statistical Tables."

31. Suppose that $f(x,t)$ is the probability of getting x successes during a time interval of length t when (i) the probability of a success during a very small time interval from t to $t + \Delta t$ is $\alpha \cdot \Delta t$, (ii) the probability of more than one success during such a time interval is negligible, and (iii) the probability of a success during such a time interval does not depend on what happened prior to time t.

(a) Show that under these conditions

$$f(x, t+\Delta t) = f(x,t)[1 - \alpha \cdot \Delta t] + f(x-1, t)\alpha \cdot \Delta t$$

and hence that

$$\frac{d[f(x,t)]}{dt} = \alpha[f(x-1,t) - f(x,t)]$$

(b) Show by direct substitution that a solution of this infinite system of differential equations (there is one for each value of x) is given by the Poisson distribution with $\lambda = \alpha t$.

32. Use repeated integration by parts to show that

$$\sum_{y=0}^{x} \frac{\lambda^y e^{-\lambda}}{y!} = \frac{1}{x!} \cdot \int_{\lambda}^{\infty} t^x e^{-t}\, dt$$

This result is important because values of the distribution function of a Poisson random variable may thus be obtained by referring to a table of incomplete gamma functions.

33. Derive the formulas for the mean and the variance of the Poisson distribution by first evaluating $E(X)$ and $E[X(X-1)]$.

34. Show that if the limiting conditions $n \to \infty, \theta \to 0$, while $n\theta$ remains constant, are applied to the moment-generating function of the binomial distribution, we get the moment-generating function of the Poisson distribution.

[*Hint*: Make use of the fact that $\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = e^z$.]

35. This question has been intentionally omitted for this edition.

36. Differentiating with respect to λ the expressions on both sides of the equation

$$\mu_r = \sum_{x=0}^{\infty} (x-\lambda)^r \cdot \frac{\lambda^x e^{-\lambda}}{x!}$$

derive the following recursion formula for the moments about the mean of the Poisson distribution:

$$\mu_{r+1} = \lambda \left[r\mu_{r-1} + \frac{d\mu_r}{d\lambda} \right]$$

for $r = 1, 2, 3, \ldots$. Also, use this recursion formula and the fact that $\mu_0 = 1$ and $\mu_1 = 0$ to find μ_2, μ_3, and μ_4, and verify the formula given for α_3 in Exercise 35.

37. Use Theorem 9 to find the moment-generating function of $Y = X - \lambda$, where X is a random variable having the Poisson distribution with the parameter λ, and use it to verify that $\sigma_Y^2 = \lambda$.

8 The Multinomial Distribution

An immediate generalization of the binomial distribution arises when each trial has more than two possible outcomes, the probabilities of the respective outcomes are the same for each trial, and the trials are all independent. This would be the case, for instance, when persons interviewed by an opinion poll are asked whether they are for a candidate, against her, or undecided or when samples of manufactured products are rated excellent, above average, average, or inferior.

To treat this kind of problem in general, let us consider the case where there are n independent trials permitting k mutually exclusive outcomes whose respective probabilities are $\theta_1, \theta_2, \ldots, \theta_k$ $\left(\text{with } \sum_{i=1}^{k} \theta_i = 1\right)$. Referring to the outcomes as being of the first kind, the second kind, ..., and the kth kind, we shall be interested in the probability of getting x_1 outcomes of the first kind, x_2 outcomes of the second kind, ..., and x_k outcomes of the kth kind $\left(\text{with } \sum_{i=1}^{k} x_i = n\right)$.

Proceeding as in the derivation of the formula for the binomial distribution, we first find that the probability of getting x_1 outcomes of the first kind, x_2 outcomes of the second kind, ..., and x_k outcomes of the kth kind *in a specific order* is $\theta_1^{x_1} \cdot \theta_2^{x_2} \cdot \ldots \cdot \theta_k^{x_k}$. To get the corresponding probability for that many outcomes of each kind *in any order*, we shall have to multiply the probability for any specific order by

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! \cdot x_2! \cdot \ldots \cdot x_k!}$$

DEFINITION 8. MULTINOMIAL DISTRIBUTION. *The random variables* $X_1, X_2, \ldots, X_n$ *have a* ***multinomial distribution*** *and they are referred to as multinomial random variables if and only if their joint probability distribution is given by*

$$f(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k) = \binom{n}{x_1, x_2, \ldots, x_k} \cdot \theta_1^{x_1} \cdot \theta_2^{x_2} \cdot \ldots \cdot \theta_k^{x_k}$$

for $x_i = 0, 1, \ldots n$ *for each* i, *where* $\sum_{i=1}^{k} x_i = n$ *and* $\sum_{i=1}^{k} \theta_i = 1$.

Thus, the numbers of outcomes of the different kinds are random variables having the multinomial distribution with the parameters $n, \theta_1, \theta_2, \ldots$, and θ_k. The name "multinomial" derives from the fact that for various values of the x_i, the probabilities equal corresponding terms of the multinomial expansion of $(\theta_1 + \theta_2 + \cdots + \theta_k)^n$.

EXAMPLE 14

A certain city has 3 newspapers, A, B, and C. Newspaper A has 50 percent of the readers in that city. Newspaper B, has 30 percent of the readers, and newspaper C has the remaining 20 percent. Find the probability that, among 8 randomly-chosen readers in that city, 5 will read newspaper A, 2 will read newspaper B, and 1 will read newspaper C. (For the purpose of this example, assume that no one reads more than one newspaper.)

Solution
Substituting $x_1 = 5$, $x_2 = 2$, $x_3 = 1$, $\theta_1 = 0.50$, $\theta_2 = 0.30$, $\theta_3 = 0.20$, and $n = 8$ into the formula of Definition 8, we get

$$\begin{aligned} f(5,2,1;8,0.50,0.30,0.20) &= \frac{8!}{5!\cdot 2!\cdot 1!}(0.50)^5(0.30)^2(0.20) \\ &= 0.0945 \end{aligned}$$

9 The Multivariate Hypergeometric Distribution

Just as the hypergeometric distribution takes the place of the binomial distribution for sampling without replacement, there also exists a multivariate distribution analogous to the multinomial distribution that applies to sampling without replacement. To derive its formula, let us consider a set of N elements, of which M_1 are elements of the first kind, M_2 are elements of the second kind, ..., and M_k are elements of the kth kind, such that $\sum_{i=1}^{k} M_i = N$. As in connection with the multinomial distribution, we are interested in the probability of getting x_1 elements (outcomes) of the first kind, x_2 elements of the second kind, ..., and x_k elements of the kth kind, but now we are choosing, without replacement, n of the N elements of the set.

There are $\binom{M_1}{x_1}$ ways of choosing x_1 of the M_1 elements of the first kind, $\binom{M_2}{x_2}$ ways of choosing x_2 of the M_2 elements of the second kind, ..., and $\binom{M_k}{x_k}$ ways of choosing x_k of the M_k elements of the kth kind, and, hence, $\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot\ldots\cdot\binom{M_k}{x_k}$ ways of choosing the required $\sum_{i=1}^{k} x_i = n$ elements. Since there are $\binom{N}{n}$ ways of choosing n of the N elements in the set and we assume that they are all equally likely (which is what we mean when we say that the selection is random), it follows that the desired probability is given by $\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot\ldots\cdot\binom{M_k}{x_k}\Big/\binom{N}{n}$.

DEFINITION 9. MULTIVARIATE HYPERGEOMETRIC DISTRIBUTION. *The random variables* $X_1, X_2, \ldots, X_k$ *have a* ***multivariate hypergeometric distribution*** *and they are referred to as multivariate hypergeometric random variables if and only if their joint probability distribution is given by*

$$f(x_1, x_2, \ldots, x_k; n, M_1, M_2, \ldots, M_k) = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot\ldots\cdot\binom{M_k}{x_k}}{\binom{N}{n}}$$

for $x_i = 0, 1, \ldots n$ *and* $x_i \le M_i$ *for each* i, *where* $\sum_{i=1}^{k} x_i = n$ *and* $\sum_{i=1}^{k} M_i = N$.

Thus, the joint distribution of the random variables under consideration, that is, the distribution of the numbers of outcomes of the different kinds, is a multivariate hypergeometric distribution with the parameters $n, M_1, M_2, \ldots$, and M_k.

EXAMPLE 15

A panel of prospective jurors includes six married men, three single men, seven married women, and four single women. If the selection is random, what is the probability that a jury will consist of four married men, one single man, five married women, and two single women?

Solution

Substituting $x_1 = 4$, $x_2 = 1$, $x_3 = 5$, $x_4 = 2$, $M_1 = 6$, $M_2 = 3, M_3 = 7$, $M_4 = 4$, $N = 20$, and $n = 12$ into the formula of Definition 9, we get

$$f(4,1,5,2;12,6,3,7,4) = \frac{\binom{6}{4}\binom{3}{1}\binom{7}{5}\binom{4}{2}}{\binom{20}{12}}$$

$$= 0.0450$$

Exercises

38. If $X_1, X_2, \ldots, X_k$ have the multinomial distribution of Definition 8, show that the mean of the marginal distribution of X_i is $n\theta_i$ for $i = 1, 2, \ldots, k$.

39. If $X_1, X_2, \ldots, X_k$ have the multinomial distribution of Definition 8, show that the covariance of X_i and X_j is $-n\theta_i\theta_j$ for $i = 1, 2, \ldots, k, j = 1, 2, \ldots, k$, and $i \neq j$.

10 The Theory in Practice

In this section we shall discuss an important application of the binomial distribution, namely **sampling inspection**.

In sampling inspection, a specified sample of a lot of manufactured product is inspected under controlled, supervised conditions. If the number of defectives found in the sample exceeds a given **acceptance number**, the lot is rejected. (A rejected lot may be subjected to closer inspection, but it is rarely scrapped.) A **sampling plan** consists of a specification of the number of items to be included in the sample taken from each lot, and a statement about the maximum number of defectives allowed before rejection takes place.

The probability that a lot will be accepted by a given sampling plan, of course, will depend upon p, the actual proportion of defectives in the lot. Since the value of p is unknown, we calculate the probability of accepting a lot for several different values of p. Suppose a sampling plan requires samples of size n from each lot, and that the lot size is large with respect to n. Suppose, further, that the acceptance number is c; that is, the lot will be accepted if c defectives or fewer are found in the sample. The probability of acceptance, the probability of finding c or fewer defectives in a sample of size n, is given by the binomial distribution to a close approximation. (Since sampling inspection is done without replacement, the assumption of equal probabilities from trial to trial, underlying the binomial distribution, is violated. But if the sample size is small relative to the lot size, this assumption is nearly satisfied.) Thus, for large lots, the probability of accepting a lot having the proportion of defectives p is closely approximated by the following definition.

> **DEFINITION 10. PROBABILITY OF ACCEPTANCE.** *If* n *is the size of the sample taken from each large lot and* c *is the acceptance number, the* ***probability of acceptance*** *is closely approximated by*
>
> $$L(p) = \sum_{k=0}^{c} b(k; n, p) = B(c; n, p)$$
>
> *where* p *is the actual proportion of defectives in the lot.*

This equation simply states that the probability of c or fewer defectives in the sample is given by the probability of 0 defectives, plus the probability of 1 defective, ..., up to the probability of c defectives, with each probability being approximated by the binomial distribution having the parameters n and $\theta = p$. Definition 10 is closely related to the power function.

It can be seen from this definition that, for a given sampling plan (sample size, n, and acceptance number, c), the probability of acceptance depends upon p, the actual (unknown) proportion of defectives in the lot. Thus a curve can be drawn that gives the probability of accepting a lot as a function of the lot proportion defective, p. This curve, called the **operating characteristic curve**, or **OC curve**, defines the characteristics of the sampling plan.

To illustrate the construction of an OC curve, let us consider the sampling plan having $n = 20$ and $c = 3$. That is, samples of size 20 are drawn from each lot, and a lot is accepted if the sample contains 3 or fewer defectives. Referring to the line in Table I of "Statistical Tables" corresponding to $n = 20$ and $x = 3$, the probabilities that a random variable having the binomial distribution $b(x; 20, p)$ will assume a value less than or equal to 3 for various values of p are as follows:

p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
$L(p)$	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049

A graph of $L(p)$ versus p is shown in Figure 6.

Inspection of the OC curve given in Figure 6 shows that the probability of acceptance is quite high (greater than 0.9) for small values of p, say values less than about 0.10. Also, the probability of acceptance is low (less than 0.10) for values of p greater than about 0.30. If the actual proportion of defectives in the lot lies between 0.10 and 0.30, however, it is somewhat of a tossup whether the lot will be accepted or rejected.

An "ideal" OC curve would be like the one shown in Figure 7. In this figure, there is no "gray area"; that is, it is certain that a lot with a given small value of p or less will be accepted, and it is certain that a lot with a value of p greater than the given value will be rejected. By comparison, the OC curve of Figure 6 seems to do a poor job of discriminating between "good" and "bad" lots. In such cases, a better OC curve can be obtained by increasing the sample size, n.

The OC curve of a sampling plan never can be like the ideal curve of Figure 7 with finite sample sizes, as there always will be some statistical error associated with sampling. However, sampling plans can be evaluated by choosing two values of p considered to be important and calculating the probabilities of lot acceptance at these values. First, a number, p_0, is chosen so that a lot containing a proportion of defectives less than or equal to p_0 is desired to be accepted. This value of p is

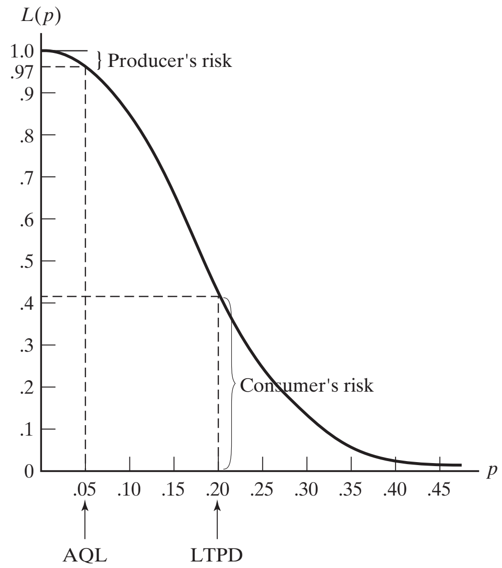

Figure 6. OC curve.

called the **acceptable quality level**, or **AQL**. Then, a second value of p, p_1, is chosen so that we wish to reject a lot containing a proportion of defectives greater than p_1. This value of p is called the **lot tolerance percentage defective**, or **LTPD**. We evaluate a sampling plan by finding the probability that a "good" lot (a lot with $p \le p_0$) will be rejected and the probability that a "bad" lot (one with $p \geq p_1$) will be accepted.

The probability that a "good" lot will be rejected is called the **producer's risk**, and the probability that a "bad" lot will be accepted is called the **consumer's risk**. The producer's risk expresses the probability that a "good" lot (one with $p < p_0$) will erroneously be rejected by the sampling plan. It is the risk that the producer takes as a consequence of sampling variability. The consumer's risk is the probability that the consumer erroneously will receive a "bad" lot (one with $p > p_1$). These risks are analogous to the type I and type II errors, α and β (If the true value of the parameter θ is θ_0 and the statistician incorrectly concludes that $\theta = \theta_1$, he is committing an error referred to as a type I error. On the other hand, if the true value of the parameter θ

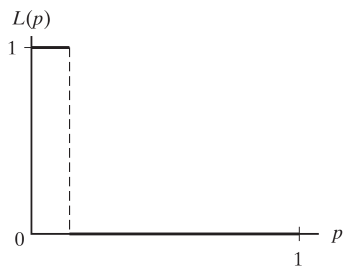

Figure 7. "Ideal" OC curve.

is θ_1 and the statistician incorrectly concludes that $\theta = \theta_0$, he is committing a type II error.)

Suppose an AQL of 0.05 is chosen ($p_0 = 0.05$). Then, it can be seen from Figure 6 that the given sampling plan has a producer's risk of about 0.03, since the probability of *acceptance* of a lot with an actual proportion defective of 0.05 is approximately 0.97. Similarly, if an LTPD of 0.20 is chosen, the consumer's risk is about 0.41. This plan obviously has an unacceptably high consumer's risk—over 40 percent of the lots received by the consumer will have 20 percent defectives or greater. To produce a plan with better characteristics, it will be necessary to increase the sample size, n, to decrease the acceptance number, c, or both. The following example shows what happens to these characteristics when c is decreased to 1, while n remains fixed at 20.

EXAMPLE 16

Find the producer's and consumer's risks corresponding to an AQL of 0.05 and an LTPD of 0.20 for the sampling plan defined by $n = 20$ and $c = 1$.

Solution

First, we calculate $L(p)$ for various values of p. Referring to Table I of "Statistical Tables" with $n = 20$ and $x = 1$, we obtain the following table:

p	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
$L(p)$	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001

A graph of this OC curve is shown in Figure 8. From this graph, we observe that the producer's risk is $1 - 0.7358 = 0.2642$, and the consumer's risk is 0.0692. Note that the work of constructing OC curves can be shortened considerably using computer software such as Excel or MINITAB.

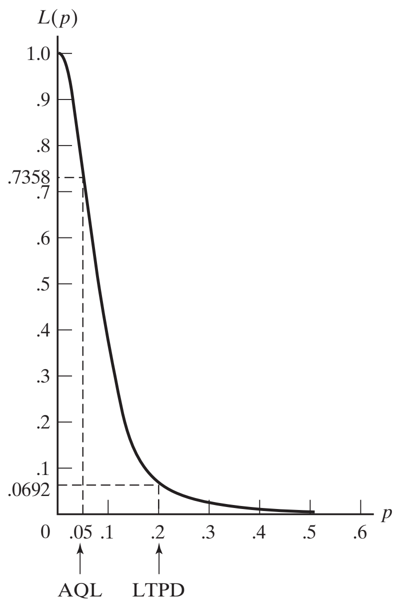

Figure 8. OC curve for Example 16.

Reduction of the acceptance number from 3 to 1 obviously has improved the consumer's risk, but now the producer's risk seems unacceptably high. Evidently, a larger sample size is needed.

The preceding example has been somewhat artificial. It would be quite unusual to specify an LTPD as high as 0.20 (20 percent defectives), and higher sample sizes than 20 usually are used for acceptance sampling. In practice, OC curves have been calculated for sampling plans having many different combinations of n and c. Choice then is made of the sampling plan whose OC curve has as nearly as possible the desired characteristics, AQL, LTPD, consumer's risk, and producer's risk for sample sizes in an acceptable range.

Applied Exercises SECS. 1–4

40. A multiple-choice test consists of eight questions and three answers to each question (of which only one is correct). If a student answers each question by rolling a balanced die and checking the first answer if he gets a 1 or 2, the second answer if he gets a 3 or 4, and the third answer if he gets a 5 or 6, what is the probability that he will get exactly four correct answers?

41. An automobile safety engineer claims that 1 in 10 automobile accidents is due to driver fatigue. Using the formula for the binomial distribution and rounding to four decimals, what is the probability that at least 3 of 5 automobile accidents are due to driver fatigue?

42. In a certain city, incompatibility is given as the legal reason in 70 percent of all divorce cases. Find the probability that five of the next six divorce cases filed in this city will claim incompatibility as the reason, using
(a) the formula for the binomial distribution;
(b) Table I of "Statistical Tables."

43. If 40 percent of the mice used in an experiment will become very aggressive within 1 minute after having been administered an experimental drug, find the probability that exactly 6 of 15 mice that have been administered the drug will become very aggressive within 1 minute, using
(a) the formula for the binomial distribution;
(b) Table I of "Statistical Tables."

44. A social scientist claims that only 50 percent of all high school seniors capable of doing college work actually go to college. Assuming that this claim is true, use Table I of "Statistical Tables" to find the probabilities that among 18 high school seniors capable of doing college work
(a) exactly 10 will go to college;
(b) at least 10 will go to college;
(c) at most 8 will go to college.

45. Suppose that the probability is 0.63 that a car stolen in a certain Western city will be recovered. Use the computer printout of Figure 1 to find the probability that at least 8 of 10 cars stolen in this city will be recovered, using
(a) the values in the P(X = K) column;
(b) the values in the P(X LESS OR = K) column.

46. With reference to Exercise 45 and the computer printout of Figure 1, find the probability that among 10 cars stolen in the given city anywhere from 3 to 5 will be recovered, using
(a) the values in the P(X = K) column;
(b) the values in the P(X LESS OR = K) column.

47. With reference to Exercise 43, suppose that the percentage had been 42 instead of 40. Use a suitable table or a computer printout of the binomial distribution with $n = 15$ and $\theta = 0.42$ to rework both parts of that exercise.

48. With reference to Exercise 44, suppose that the percentage had been 51 instead of 50. Use a suitable table or a computer printout of the binomial distribution with $n = 18$ and $\theta = 0.51$ to rework the three parts of that exercise.

49. In planning the operation of a new school, one school board member claims that four out of five newly hired teachers will stay with the school for more than a year, while another school board member claims that it would be correct to say three out of five. In the past, the two board members have been about equally reliable in their predictions, so in the absence of any other information we would assign their judgments equal weight. If one or the other has to be right, what probabilities would we assign to their claims if it were found that 11 of 12 newly hired teachers stayed with the school for more than a year?

50. (a) To reduce the standard deviation of the binomial distribution by half, what change must be made in the number of trials?
(b) If n is multiplied by the factor k in the binomial distribution having the parameters n and θ, what statement can be made about the standard deviation of the resulting distribution?

51. A manufacturer claims that at most 5 percent of the time a given product will sustain fewer than 1,000 hours of operation before requiring service. Twenty products were selected at random from the production line and tested. It was found that three of them required service before 1,000 hours of operation. Comment on the manufacturer's claim.

52. (a) Use a computer program to calculate the probability of rolling between 14 and 18 "sevens" in 100 rolls of a pair of dice.

(b) Would it surprise you if more than 18 "sevens" were rolled? Why?

53. (a) Use a computer program to calculate the probability that more than 12 of 80 business telephone calls last longer than five minutes if it is assumed that 10 percent of such calls last that long.

(b) Can this result be used as evidence that the assumption is reasonable? Why?

54. Use Chebyshev's theorem and Theorem 3 to verify that the probability is at least $\frac{35}{36}$ that

(a) in 900 flips of a balanced coin the proportion of heads will be between 0.40 and 0.60;

(b) in 10,000 flips of a balanced coin the proportion of heads will be between 0.47 and 0.53;

(c) in 1,000,000 flips of a balanced coin the proportion of heads will be between 0.497 and 0.503.

Note that this serves to illustrate the law of large numbers.

55. You can get a feeling for the law of large numbers given Section 4 by flipping coins. Flip a coin 100 times and plot the accumulated proportion of heads after each five flips.

56. Record the first 200 numbers encountered in a newspaper, beginning with page 1 and proceeding in any convenient, systematic fashion. Include also numbers appearing in advertisements. For each of these numbers, note the leftmost digit, and record the proportions of 1's, 2's, 3's, . . . , and 9's. (Note that 0 cannot be a leftmost digit. In the decimal number 0.0074, the leftmost digit is 7.) The results may seem quite surprising, but the law of large numbers tells you that you must be estimating correctly.

SECS. 5–7

57. If the probabilities of having a male or female child are both 0.50, find the probabilities that

(a) a family's fourth child is their first son;

(b) a family's seventh child is their second daughter;

(c) a family's tenth child is their fourth or fifth son.

58. If the probability is 0.75 that a person will believe a rumor about the transgressions of a certain politician, find the probabilities that

(a) the eighth person to hear the rumor will be the fifth to believe it;

(b) the fifteenth person to hear the rumor will be the tenth to believe it.

59. When taping a television commercial, the probability is 0.30 that a certain actor will get his lines straight on any one take. What is the probability that he will get his lines straight for the first time on the sixth take?

60. An expert sharpshooter misses a target 5 percent of the time. Find the probability that she will miss the target for the second time on the fifteenth shot using

(a) the formula for the negative binomial distribution;

(b) Theorem 5 and Table I of "Statistical Tables."

61. Adapt the formula of Theorem 5 so that it can be used to express geometric probabilities in terms of binomial probabilities, and use the formula and Table I of "Statistical Tables" to

(a) verify the result of Example 5;

(b) rework Exercise 59.

62. In a "torture test" a light switch is turned on and off until it fails. If the probability is 0.001 that the switch will fail any time it is turned on or off, what is the probability that the switch will not fail during the first 800 times that it is turned on or off? Assume that the conditions underlying the geometric distribution are met and use logarithms.

63. A quality control engineer inspects a random sample of two hand-held calculators from each incoming lot of size 18 and accepts the lot if they are both in good working condition; otherwise, the entire lot is inspected with the cost charged to the vendor. What are the probabilities that such a lot will be accepted without further inspection if it contains

(a) 4 calculators that are not in good working condition;

(b) 8 calculators that are not in good working condition;

(c) 12 calculators that are not in good working condition?

64. Among the 16 applicants for a job, 10 have college degrees. If 3 of the applicants are randomly chosen for interviews, what are the probabilities that

(a) none has a college degree;

(b) 1 has a college degree;

(c) 2 have college degrees;

(d) all 3 have college degrees?

65. Find the mean and the variance of the hypergeometric distribution with $n = 3, N = 16$, and $M = 10$, using

(a) the results of Exercise 64;

(b) the formulas of Theorem 7.

66. What is the probability that an IRS auditor will catch only 2 income tax returns with illegitimate deductions if she randomly selects 5 returns from among 15 returns, of which 9 contain illegitimate deductions?

67. Check in each case whether the condition for the binomial approximation to the hypergeometric distribution is satisfied:

(a) $N = 200$ and $n = 12$;

(b) $N = 500$ and $n = 20$;

(c) $N = 640$ and $n = 30$.

68. A shipment of 80 burglar alarms contains 4 that are defective. If 3 from the shipment are randomly selected and shipped to a customer, find the probability that the customer will get exactly one bad unit using

(a) the formula of the hypergeometric distribution;

(b) the binomial distribution as an approximation.

69. Among the 300 employees of a company, 240 are union members, whereas the others are not. If 6 of the employees are chosen by lot to serve on a committee that administers the pension fund, find the probability that 4 of the 6 will be union members using

(a) the formula for the hypergeometric distribution;

(b) the binomial distribution as an approximation.

70. A panel of 300 persons chosen for jury duty includes 30 under 25 years of age. Since the jury of 12 persons chosen from this panel to judge a narcotics violation does not include anyone under 25 years of age, the youthful defendant's attorney complains that this jury is not really representative. Indeed, he argues, if the selection were random, the probability of having one of the 12 jurors under 25 years of age should be *many times* the probability of having none of them under 25 years of age. Actually, what is the ratio of these two probabilities?

71. Check in each case whether the values of n and θ satisfy the rule of thumb for a good approximation, an excellent approximation, or neither when we want to use the Poisson distribution to approximate binomial probabilities.

(a) $n = 125$ and $\theta = 0.10$;

(b) $n = 25$ and $\theta = 0.04$;

(c) $n = 120$ and $\theta = 0.05$;

(d) $n = 40$ and $\theta = 0.06$.

72. It is known from experience that 1.4 percent of the calls received by a switchboard are wrong numbers. Use the Poisson approximation to the binomial distribution to determine the probability that among 150 calls received by the switchboard 2 are wrong numbers.

73. With reference to Example 8, determine the value of x (from 5 to 15) for which the percentage error is greatest when we use the Poisson distribution with $\lambda = 7.5$ to approximate the binomial distribution with $n = 150$ and $\theta = 0.05$.

74. In a given city, 4 percent of all licensed drivers will be involved in at least one car accident in any given year. Use the Poisson approximation to the binomial distribution to determine the probability that among 150 licensed drivers randomly chosen in this city

(a) only 5 will be involved in at least one accident in any given year;

(b) at most 3 will be involved in at least one accident in any given year.

75. Records show that the probability is 0.0012 that a person will get food poisoning spending a day at a certain state fair. Use the Poisson approximation to the binomial distribution to find the probability that among 1,000 persons attending the fair at most 2 will get food poisoning.

76. With reference to Example 13 and the computer printout of Figure 4, find the probability that a 15-square-foot sheet of the metal will have anywhere from 8 to 12 defects, using

(a) the values in the P(X = K) column;

(b) the values in the P(X LESS OR = K) column.

77. The number of complaints that a dry-cleaning establishment receives per day is a random variable having a Poisson distribution with $\lambda = 3.3$. Use the formula for the Poisson distribution to find the probability that it will receive only two complaints on any given day.

78. The number of monthly breakdowns of a super computer is a random variable having a Poisson distribution with $\lambda = 1.8$. Use the formula for the Poisson distribution to find the probabilities that this computer will function

(a) without a breakdown;

(b) with only one breakdown.

79. Use Table II of "Statistical Tables" to verify the results of Exercise 78.

80. In the inspection of a fabric produced in continuous rolls, the number of imperfections per yard is a random variable having the Poisson distribution with $\lambda = 0.25$. Find the probability that 2 yards of the fabric will have at most one imperfection using

(a) Table II of "Statistical Tables";

(b) the computer printout of Figure 5.

81. In a certain desert region the number of persons who become seriously ill each year from eating a certain poisonous plant is a random variable having a Poisson distribution with $\lambda = 5.2$. Use Table II of "Statistical Tables" to find the probabilities of

(a) 3 such illnesses in a given year;

(b) at least 10 such illnesses in a given year;

(c) anywhere from 4 to 6 such illnesses in a given year.

82. (a) Use a computer program to calculate the *exact* probability of obtaining one or more defectives in a sample of size 100 taken from a lot of 1,000 manufactured products assumed to contain six defectives.

(b) Approximate this probability using the appropriate binomial distribution.

(c) Approximate this probability using the appropriate Poisson distribution and compare the results of parts (a), (b), and (c).

SECS. 8–9

83. The probabilities are 0.40, 0.50, and 0.10 that, in city driving, a certain kind of compact car will average less than 28 miles per gallon, from 28 to 32 miles per gallon, or more than 32 miles per gallon. Find the probability that among 10 such cars tested, 3 will average less than

28 miles per gallon, 6 will average from 28 to 32 miles per gallon, and 1 will average more than 32 miles per gallon.

84. Suppose that the probabilities are 0.60, 0.20, 0.10, and 0.10 that a state income tax return will be filled out correctly, that it will contain only errors favoring the taxpayer, that it will contain only errors favoring the state, or that it will contain both kinds of errors. What is the probability that among 12 such income tax returns randomly chosen for audit, 5 will be filled out correctly, 4 will contain only errors favoring the taxpayer, 2 will contain only errors favoring the state, and 1 will contain both kinds of errors?

85. According to the Mendelian theory of heredity, if plants with round yellow seeds are crossbred with plants with wrinkled green seeds, the probabilities of getting a plant that produces round yellow seeds, wrinkled yellow seeds, round green seeds, or wrinkled green seeds are, respectively, $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$. What is the probability that among nine plants thus obtained there will be four that produce round yellow seeds, two that produce wrinkled yellow seeds, three that produce round green seeds, and none that produce wrinkled green seeds?

86. Among 25 silver dollars struck in 1903 there are 15 from the Philadelphia mint, 7 from the New Orleans mint, and 3 from the San Francisco mint. If 5 of these silver dollars are picked at random, find the probabilities of getting

(a) 4 from the Philadelphia mint and 1 from the New Orleans mint;

(b) 3 from the Philadelphia mint and 1 from each of the other 2 mints.

87. If 18 defective glass bricks include 10 that have cracks but no discoloration, 5 that have discoloration but no cracks, and 3 that have cracks and discoloration, what is the probability that among 6 of the bricks (chosen at random for further checks) 3 will have cracks but no discoloration, 1 will have discoloration but no cracks, and 2 will have cracks and discoloration?

SEC. 10

88. A sampling inspection program has a 0.10 probability of rejecting a lot when the true proportion of defectivesis 0.01, and a 0.95 probability of rejecting the lot when the true proportion of defectives is 0.03. If 0.01 is the AQL and 0.03 is the LTPD, what are the producer's and consumer's risks?

89. The producer's risk in a sampling program is 0.05 and the consumer's risk is 0.10; the AQL is 0.03 and the LTPD is 0.07.

(a) What is the probability of accepting a lot whose true proportion of defectives is 0.03?

(b) What is the probability of accepting a lot whose true proportion of defectives is 0.07?

90. Suppose the acceptance number in Example 16 is changed from 1 to 2. Keeping the producer's risk at 0.05 and the consumer's risk at 0.10, what are the new values of the AQL and the LTPD?

91. From Figure 6,

(a) find the producer's risk if the AQL is 0.10;

(b) find the LTPD corresponding to a consumer's risk of 0.05.

92. Sketch the OC curve for a sampling plan having a sample size of 15 and an acceptance number of 1.

93. Sketch the OC curve for a sampling plan having a sample size of 25 and an acceptance number of 2.

94. Sketch the OC curve for a sampling plan having a sample size of 10 and an acceptance number of 0.

95. Find the AQL and the LTPD of the sampling plan in Exercise 93 if both the producer's and consumer's risks are 0.10.

96. If the AQL is 0.1 and the LTPD is 0.25 in the sampling plan given in Exercise 92, find the producer's and consumer's risks.

97. (a) In Exercise 92 change the acceptance number from 1 to 0 and sketch the OC curve.

(b) How do the producer's and consumer's risks change if the AQL is 0.05 and the LTPD is 0.3 in both sampling plans?

References

Useful information about various special probability distributions may be found in

Derman, C., Gleser, L., and Olkin, I., *Probability Models and Applications*. New York: Macmillan Publishing Co., Inc., 1980,

Hastings, N. A. J., and Peacock, J. B., *Statistical Distributions*. London: Butterworth & Co. Ltd., 1975,

and

Johnson, N. L., and Kotz, S., *Discrete Distributions*, Boston: Houghton Mifflin Company, 1969.

Binomial probabilities for $n = 2$ to $n = 49$ may be found in

Tables of the Binomial Probability Distribution, National Bureau of Standards Applied Mathematics Series No. 6, Washington, D.C.: U.S. Government Printing Office, 1950, and for $n = 50$ to $n = 100$ in

ROMIG, H. G., 50–100 *Binomial Tables*. New York: John Wiley & Sons, Inc., 1953.

The most widely used table of Poisson probabilities is

MOLINA, E. C., *Poisson's Exponential Binomial Limit*. Melbourne, Fla.: Robert E. Krieger Publishing Company, 1973 Reprint.

Answers to Odd-Numbered Exercises

11 $\mu_2' = \mu_{(2)}' + \mu_{(1)}'$, $\mu_3' = \mu_{(3)}' + 3\mu_{(2)}' + \mu_{(1)}'$, and $\mu_4' = \mu_{(4)}' + 6\mu_{(3)}' + 7\mu_{(2)}' + \mu_{(1)}'$.

13 **(a)** $F_X(t) = 1 - \theta + \theta t$; **(b)** $F_X(t) = [1 + \theta(t-1)]^n$.

15 **(a)** $\alpha_3 = 0$ when $\theta = \frac{1}{2}$; **(b)** $\alpha_3 \to 0$ when $n \to \infty$.

17 $\mu_Y = k\left(\frac{1}{\theta} - 1\right)$; $\sigma_Y^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right)$.

37 $M_Y(t) = e^{\lambda(e^t - t - 1)}$; $\sigma_Y^2 = M_Y'(0) = \lambda$.

41 0.0086.

43 **(a)** 0.2066; **(b)** 0.2066.

45 **(a)** 0.2205; **(b)** 0.2206.

47 0.2041.

49 0.9222.

51 0.0754.

53 **(a)** 0.0538.

57 **(a)** 0.0625; **(b)** 0.0469; **(c)** 0.2051.

59 0.0504.

61 **(a)** 0.0117; **(b)** 0.0504.

63 **(a)** 0.5948; **(b)** 0.2941; **(c)** 0.0980.

65 **(a)** $\mu = \frac{15}{8}$ and $\sigma^2 = \frac{39}{64}$; **(b)** $\mu = \frac{15}{8}$ and $\sigma^2 = \frac{39}{64}$.

67 **(a)** The condition is not satisfied. **(b)** The condition is satisfied. **(c)** The condition is satisfied.

69 **(a)** 0.2478; **(b)** 0.2458.

71 **(a)** Neither rule of thumb is satisfied. **(b)** The rule of thumb for good approximation is satisfied. **(c)** The rule of thumb for excellent approximation is satisfied. **(d)** Neither rule of thumb is satisfied.

73 $x = 15$.

75 0.8795.

77 0.2008.

79 **(a)** 0.1653; **(b)** 0.2975.

81 **(a)** 0.1293; **(b)** 0.0397; **(c)** 0.4944.

83 0.0841.

85 0.0292.

87 0.0970.

89 **(a)** 0.95; **(b)** 0.10.

91 **(a)** 0.17; **(b)** 0.35.

95 AQL = 0.07, LTPD = 0.33.

97 **(b)** Plan 1 ($c = 0$): producer's risk = 0.0861 and consumer's risk = 0.1493; Plan 2 ($c = 1$): producer's risk = 0.4013 and consumer's risk = 0.0282.

Special Probability Densities

1 Introduction

In this chapter we shall study some of the probability densities that figure most prominently in statistical theory and in applications. In addition to the ones given in the text, several others are introduced in the exercises following Section 4. We shall derive parameters and moment-generating functions, again leaving some of the details as exercises.

2 The Uniform Distribution

Definition 1. Uniform Distribution. *A random variable* X *has a* ***uniform distribution*** *and it is referred to as a continuous uniform random variable if and only if its probability density is given by*

$$u(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{for } \alpha < x < \beta \\ 0 & \text{elsewhere} \end{cases}$$

The parameters α and β of this probability density are real constants, with $\alpha < \beta$, and may be pictured as in Figure 1. In Exercise 2 the reader will be asked to verify the following theorem.

Theorem 1. The mean and the variance of the uniform distribution are given by

$$\mu = \frac{\alpha+\beta}{2} \quad \text{and} \quad \sigma^2 = \frac{1}{12}(\beta-\alpha)^2$$

From Chapter 6 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

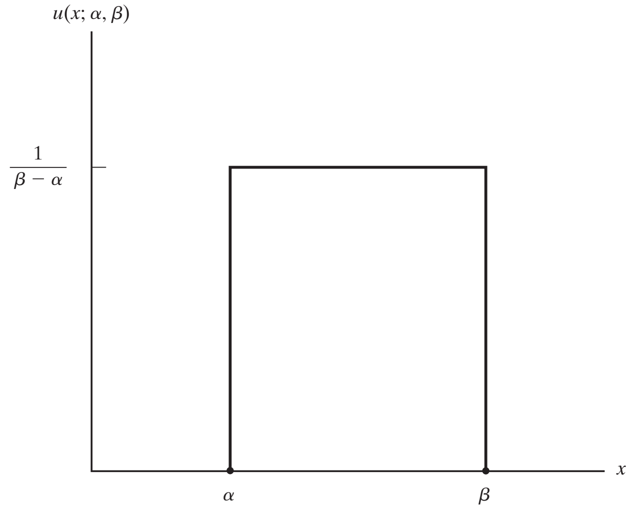

Figure 1. The uniform distribution.

Although the uniform distribution has some direct applications, its main value is that, due to its simplicity, it lends itself readily to the task of illustrating various aspects of statistical theory.

3 The Gamma, Exponential, and Chi-Square Distributions

Let's start with random variables having probability densities of the form

$$f(x) = \begin{cases} kx^{\alpha-1}e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$, $\beta > 0$, and k must be such that the total area under the curve is equal to 1. To evaluate k, we first make the substitution $y = \frac{x}{\beta}$, which yields

$$\int_0^\infty kx^{\alpha-1}e^{-x/\beta}dx = k\beta^\alpha \int_0^\infty y^{\alpha-1}e^{-y}dy$$

The integral thus obtained depends on α alone, and it defines the well-known **gamma function**

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy \qquad \text{for } \alpha > 0$$

which is treated in detail in most advanced calculus texts. Integrating by parts, which is left to the reader in Exercise 7, we find that the gamma function satisfies the recursion formula

$$\Gamma(\alpha) = (\alpha-1)\cdot\Gamma(\alpha-1)$$

for $\alpha > 1$, and since

$$\Gamma(1) = \int_0^\infty e^{-y}\,dy = 1$$

it follows by repeated application of the recursion formula that $\Gamma(\alpha) = (\alpha - 1)!$ when α is a positive integer. Also, an important special value is $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, as the reader will be asked to verify in Exercise 9.

Returning now to the problem of evaluating k, we equate the integral we obtained to 1, getting

$$\int_0^\infty kx^{\alpha-1}e^{-x/\beta}\,dx = k\beta^\alpha\Gamma(\alpha) = 1$$

and hence

$$k = \frac{1}{\beta^\alpha\Gamma(\alpha)}$$

This leads to the following definition of the **gamma distribution**.

DEFINITION 2. GAMMA DISTRIBUTION. *A random variable X has a **gamma distribution** and it is referred to as a gamma random variable if and only if its probability density is given by*

$$g(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ *and* $\beta > 0$.

When α is not a positive integer, the value of $\Gamma(\alpha)$ will have to be looked up in a special table. To give the reader some idea about the shape of the graphs of gamma densities, those for several special values of α and β are shown in Figure 2.

Some special cases of the gamma distribution play important roles in statistics; for instance, for $\alpha = 1$ and $\beta = \theta$, we obtain the following definition.

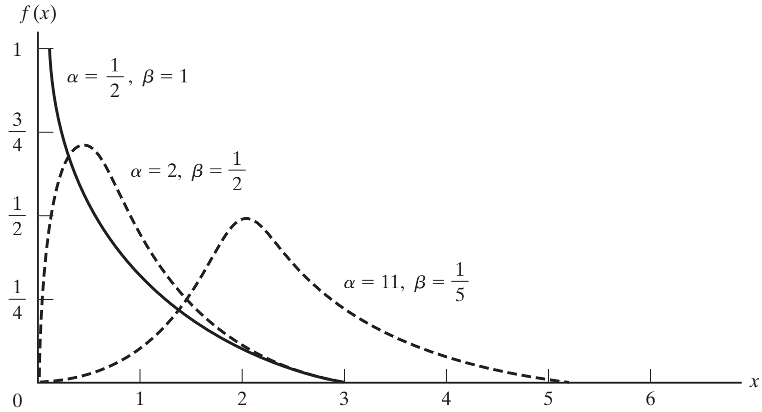

Figure 2. Graphs of gamma distributions.

DEFINITION 3. EXPONENTIAL DISTRIBUTION. *A random variable* X *has an* ***exponential distribution*** *and it is referred to as an exponential random variable if and only if its probability density is given by*

$$g(x;\theta) = \begin{cases} \dfrac{1}{\theta}e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\theta > 0$.

This density is pictured in Figure 3.

Let us consider there is the probability of getting x successes during a time interval of length t when (i) the probability of a success during a very small time interval from t to $t + \Delta t$ is $\alpha \cdot \Delta t$, (ii) the probability of more than one success during such a time interval is negligible, and (iii) the probability of a success during such a time interval does not depend on what happened prior to time t. The number of successes is a value of the discrete random variable X having the Poisson distribution with $\lambda = \alpha t$. Let us determine the probability density of the continuous random variable Y, the **waiting time** until the first success. Clearly,

$$\begin{aligned} F(y) = P(Y \leqq y) &= 1 - P(Y > y) \\ &= 1 - P(0 \text{ successes in a time interval of length } y) \\ &= 1 - p(0; \alpha y) \\ &= 1 - \frac{e^{-\alpha y}(\alpha y)^0}{0!} \\ &= 1 - e^{-\alpha y} \quad \text{for } y > 0 \end{aligned}$$

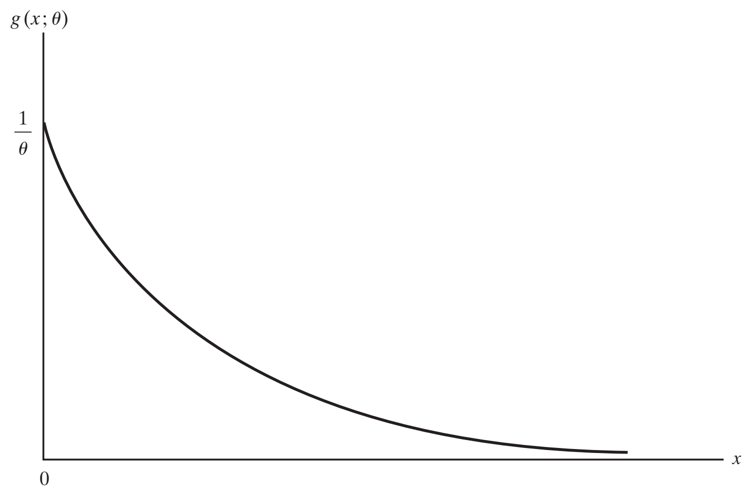

Figure 3. Exponential distribution.

and $F(y) = 0$ for $y \leqq 0$. Having thus found the distribution function of Y, we find that differentiation with respect to y yields

$$f(y) = \begin{cases} \alpha e^{-\alpha y} & \text{for } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

which is the exponential distribution with $\theta = \dfrac{1}{\alpha}$.

The exponential distribution applies not only to the occurrence of the first success in a **Poisson process** but, by virtue of condition (see Exercise 16), it applies also to the waiting times between successes.

EXAMPLE 1

At a certain location on highway I-10, the number of cars exceeding the speed limit by more than 10 miles per hour in half an hour is a random variable having a Poisson distribution with $\lambda = 8.4$. What is the probability of a waiting time of less than 5 minutes between cars exceeding the speed limit by more than 10 miles per hour?

Solution

Using half an hour as the unit of time, we have $\alpha = \lambda = 8.4$. Therefore, the waiting time is a random variable having an exponential distribution with $\theta = \frac{1}{8.4}$, and since 5 minutes is $\frac{1}{6}$ of the unit of time, we find that the desired probability is

$$\int_0^{1/6} 8.4e^{-8.4x}\,dx = -e^{-8.4x}\Big|_0^{1/6} = -e^{-1.4} + 1$$

which is approximately 0.75.

Another special case of the gamma distribution arises when $\alpha = \dfrac{\nu}{2}$ and $\beta = 2$, where ν is the lowercase Greek letter *nu*.

DEFINITION 4. CHI-SQUARE DISTRIBUTION. *A random variable* X *has a* ***chi-square distribution*** *and it is referred to as a chi-square random variable if and only if its probability density is given by*

$$f(x, \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The parameter ν is referred to as the **number of degrees of freedom**, or simply the **degrees of freedom**. The chi-square distribution plays a very important role in sampling theory.

To derive formulas for the mean and the variance of the gamma distribution, and hence the exponential and chi-square distributions, let us first prove the following theorem.

THEOREM 2. The rth moment about the origin of the gamma distribution is given by

$$\mu'_r = \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$$

Proof By using the definition of the rth moment about the origin,

$$\mu'_r = \int_0^\infty x^r \cdot \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx = \frac{\beta^r}{\Gamma(\alpha)} \cdot \int_0^\infty y^{\alpha+r-1} e^{-y} dy$$

where we let $y = \frac{x}{\beta}$. Since the integral on the right is $\Gamma(r + \alpha)$ according to the definition of gamma function, this completes the proof.

Using this theorem, let us now derive the following results about the gamma distribution.

THEOREM 3. The mean and the variance of the gamma distribution are given by

$$\mu = \alpha\beta \quad \text{and} \quad \sigma^2 = \alpha\beta^2$$

Proof From Theorem 2 with $r = 1$ and $r = 2$, we get

$$\mu'_1 = \frac{\beta\Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha\beta$$

and

$$\mu'_2 = \frac{\beta^2\Gamma(\alpha + 2)}{\Gamma(\alpha)} = \alpha(\alpha + 1)\beta^2$$

so $\mu = \alpha\beta$ and $\sigma^2 = \alpha(\alpha + 1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2$.

Substituting into these formulas $\alpha = 1$ and $\beta = \theta$ for the exponential distribution and $\alpha = \frac{\nu}{2}$ and $\beta = 2$ for the chi-square distribution, we obtain the following corollaries.

COROLLARY 1. The mean and the variance of the exponential distribution are given by

$$\mu = \theta \quad \text{and} \quad \sigma^2 = \theta^2$$

COROLLARY 2. The mean and the variance of the chi-square distribution are given by

$$\mu = \nu \quad \text{and} \quad \sigma^2 = 2\nu$$

For future reference, let us give here also the moment-generating function of the gamma distribution.

THEOREM 4. The moment-generating function of the gamma distribution is given by

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

The reader will be asked to prove this result and use it to find some of the lower moments in Exercises 12 and 13.

4 The Beta Distribution

The uniform density $f(x) = 1$ for $0 < x < 1$ and $f(x) = 0$ elsewhere is a special case of the **beta distribution**, which is defined in the following way.

DEFINITION 5. BETA DISTRIBUTION. *A random variable* X *has a* ***beta distribution*** *and it is referred to as a beta random variable if and only if its probability density is given by*

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ *and* $\beta > 0$.

In recent years, the beta distribution has found important applications in **Bayesian inference**, where parameters are looked upon as random variables, and there is a need for a fairly "flexible" probability density for the parameter θ of the binomial distribution, which takes on nonzero values only on the interval from 0 to 1. By "flexible" we mean that the probability density can take on a great variety of different shapes, as the reader will be asked to verify for the beta distribution in Exercise 27.

We shall not prove here that the total area under the curve of the beta distribution, like that of any probability density, is equal to 1, but in the proof of the theorem that follows, we shall make use of the fact that

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = 1$$

and hence that

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

This integral defines the **beta function**, whose values are denoted $B(\alpha, \beta)$; in other words, $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Detailed discussion of the beta function may be found in any textbook on advanced calculus.

THEOREM 5. The mean and the variance of the beta distribution are given by

$$\mu = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Proof By definition,

$$\begin{aligned}\mu &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\int_0^1 x\cdot x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+1)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\ &= \frac{\alpha}{\alpha+\beta}\end{aligned}$$

where we recognized the integral as $\mathrm{B}(\alpha+1,\beta)$ and made use of the fact that $\Gamma(\alpha+1) = \alpha\cdot\Gamma(\alpha)$ and $\Gamma(\alpha+\beta+1) = (\alpha+\beta)\cdot\Gamma(\alpha+\beta)$. Similar steps, which will be left to the reader in Exercise 28, yield

$$\mu_2' = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}$$

and it follows that

$$\begin{aligned}\sigma^2 &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\end{aligned}$$

Exercises

1. Show that if a random variable has a uniform density with the parameters α and β, the probability that it will take on a value less than $\alpha+p(\beta-\alpha)$ is equal to p.

2. Prove Theorem 1.

3. If a random variable X has a uniform density with the parameters α and β, find its distribution function.

4. Show that if a random variable has a uniform density with the parameters α and β, the rth moment about the mean equals
(a) 0 when r is odd;
(b) $\dfrac{1}{r+1}\left(\dfrac{\beta-\alpha}{2}\right)^r$ when r is even.

5. Use the results of Exercise 4 to find α_3 and α_4 for the uniform density with the parameters α and β.

6. A random variable is said to have a **Cauchy distribution** if its density is given by

$$f(x) = \frac{\frac{\beta}{\pi}}{(x-\alpha)^2+\beta^2} \qquad \text{for } -\infty < x < \infty$$

Show that for this distribution μ_1' and μ_2' do not exist.

7. Use integration by parts to show that $\Gamma(\alpha) = (\alpha-1)\cdot\Gamma(\alpha-1)$ for $\alpha > 1$.

8. Perform a suitable change of variable to show that the integral defining the gamma function can be written as

$$\Gamma(\alpha) = 2^{1-\alpha}\cdot\int_0^\infty z^{2\alpha-1}e^{-\frac{1}{2}z^2}dz \qquad \text{for } \alpha > 0$$

9. Using the form of the gamma function of Exercise 8, we can write

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}z^2}dz$$

and hence

$$\begin{aligned}\left[\Gamma\left(\frac{1}{2}\right)\right]^2 &= 2\left\{\int_0^\infty e^{-\frac{1}{2}x^2}dx\right\}\left\{\int_0^\infty e^{-\frac{1}{2}y^2}dy\right\} \\ &= 2\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy\end{aligned}$$

Change to polar coordinates to evaluate this double integral, and thus show that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

10. Find the probabilities that the value of a random variable will exceed 4 if it has a gamma distribution with
(a) $\alpha = 2$ and $\beta = 3$;
(b) $\alpha = 3$ and $\beta = 4$.

11. Show that a gamma distribution with $\alpha > 1$ has a relative maximum at $x = \beta(\alpha - 1)$. What happens when $0 < \alpha < 1$ and when $\alpha = 1$?

12. Prove Theorem 4, making the substitution $y = x\left(\frac{1}{\beta} - t\right)$ in the integral defining $M_X(t)$.

13. Expand the moment-generating function of the gamma distribution as a binomial series, and read off the values of μ_1', μ_2', μ_3', and μ_4'.

14. Use the results of Exercise 13 to find α_3 and α_4 for the gamma distribution.

15. Show that if a random variable has an exponential density with the parameter θ, the probability that it will take on a value less than $-\theta \cdot \ln(1-p)$ is equal to p for $0 \leqq p < 1$.

16. If X has an exponential distribution, show that

$$P[(X \geq t + T)|(x \geq T)] = P(X \geq t)$$

17. This question has been intentionally omitted for this edition.

18. With reference to Exercise 17, using the fact that the moments of Y about the origin are the corresponding moments of X about the mean, find α_3 and α_4 for the exponential distribution with the parameter θ.

19. Show that if $\nu > 2$, the chi-square distribution has a relative maximum at $x = \nu - 2$. What happens when $\nu = 2$ or $0 < \nu < 2$?

20. A random variable X has a **Rayleigh distribution** if and only if its probability density is given by

$$f(x) = \begin{cases} 2\alpha x e^{-\alpha x^2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$. Show that for this distribution

(a) $\mu = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$;

(b) $\sigma^2 = \frac{1}{\alpha}\left(1 - \frac{\pi}{4}\right)$.

21. A random variable X has a **Pareto distribution** if and only if its probability density is given by

$$f(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} & \text{for } x > 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$. Show that μ_r' exists only if $r < \alpha$.

22. With reference to Exercise 21, show that for the Pareto distribution

$$\mu = \frac{\alpha}{\alpha - 1} \text{ provided } \alpha > 1.$$

23. A random variable X has a **Weibull distribution** if and only if its probability density is given by

$$f(x) = \begin{cases} kx^{\beta-1}e^{-\alpha x^{\beta}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

(a) Express k in terms of α and β.

(b) Show that $\mu = \alpha^{-1/\beta}\Gamma\left(1 + \frac{1}{\beta}\right)$.

Note that Weibull distributions with $\beta = 1$ are exponential distributions.

24. If the random variable T is the time to failure of a commercial product and the values of its probability density and distribution function at time t are $f(t)$ and $F(t)$, then its failure rate at time t is given by $\frac{f(t)}{1 - F(t)}$. Thus, the failure rate at time t is the probability density of failure at time t given that failure does not occur prior to time t.

(a) Show that if T has an exponential distribution, the failure rate is constant.

(b) Show that if T has a Weibull distribution (see Exercise 23), the failure rate is given by $\alpha\beta t^{\beta-1}$.

25. Verify that the integral of the beta density from $-\infty$ to ∞ equals 1 for

(a) $\alpha = 2$ and $\beta = 4$;

(b) $\alpha = 3$ and $\beta = 3$.

26. Show that if $\alpha > 1$ and $\beta > 1$, the beta density has a relative maximum at

$x = \frac{\alpha - 1}{\alpha + \beta - 2}$.

27. Sketch the graphs of the beta densities having

(a) $\alpha = 2$ and $\beta = 2$;

(b) $\alpha = \frac{1}{2}$ and $\beta = 1$;

(c) $\alpha = 2$ and $\beta = \frac{1}{2}$;

(d) $\alpha = 2$ and $\beta = 5$.

[*Hint*: To evaluate $\Gamma(\frac{3}{2})$ and $\Gamma(\frac{5}{2})$, make use of the recursion formula $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ and the result of Exercise 9.]

28. Verify the expression given for μ_2' in the proof of Theorem 5.

29. Show that the parameters of the beta distribution can be expressed as follows in terms of the mean and the variance of this distribution:

(a) $\alpha = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$;

(b) $\beta = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$.

30. Karl Pearson, one of the founders of modern statistics, showed that the differential equation

$$\frac{1}{f(x)} \cdot \frac{d[f(x)]}{dx} = \frac{d - x}{a + bx + cx^2}$$

yields (for appropriate values of the constants a, b, c, and d) most of the important distributions of statistics. Verify that the differential equation gives
(a) the gamma distribution when $a = c = 0$, $b > 0$, and $d > -b$;
(b) the exponential distribution when $a = c = d = 0$ and $b > 0$;
(c) the beta distribution when $a = 0$, $b = -c$, $\frac{d-1}{b} < 1$, and $\frac{d}{b} > -1$.

5 The Normal Distribution

The **normal distribution**, which we shall study in this section, is in many ways the cornerstone of modern statistical theory. It was investigated first in the eighteenth century when scientists observed an astonishing degree of regularity in errors of measurement. They found that the patterns (distributions) that they observed could be closely approximated by continuous curves, which they referred to as "normal curves of errors" and attributed to the laws of chance. The mathematical properties of such normal curves were first studied by Abraham de Moivre (1667–1745), Pierre Laplace (1749–1827), and Karl Gauss (1777–1855).

DEFINITION 6. NORMAL DISTRIBUTION. *A random variable* X *has a* ***normal distribution*** *and it is referred to as a normal random variable if and only if its probability density is given by*

$$n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

where $\sigma > 0$.

The graph of a normal distribution, shaped like the cross section of a bell, is shown in Figure 4.

The notation used here shows explicitly that the two parameters of the normal distribution are μ and σ. It remains to be shown, however, that the parameter μ is, in fact, $E(X)$ and that the parameter σ is, in fact, the square root of $\text{var}(X)$, where X is a random variable having the normal distribution with these two parameters.

First, though, let us show that the formula of Definition 6 can serve as a probability density. Since the values of $n(x;\mu,\sigma)$ are evidently positive as long as $\sigma > 0$,

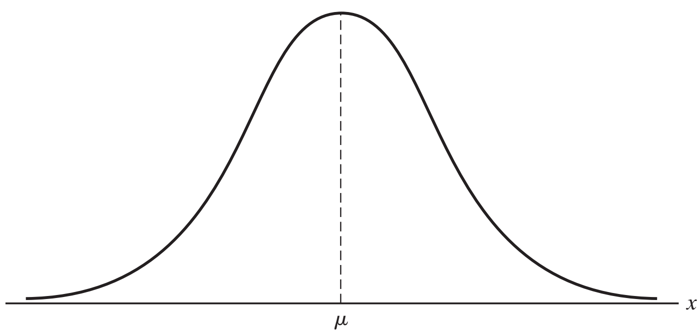

Figure 4. Graph of normal distribution.

we must show that the total area under the curve is equal to 1. Integrating from $-\infty$ to ∞ and making the substitution $z = \dfrac{x-\mu}{\sigma}$, we get

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}z^2} dz$$

Then, since the integral on the right equals $\dfrac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{2}} = \dfrac{\sqrt{\pi}}{\sqrt{2}}$ according to Exercise 9, it follows that the total area under the curve is equal to $\dfrac{2}{\sqrt{2\pi}} \cdot \dfrac{\sqrt{\pi}}{\sqrt{2}} = 1$.

Next let us prove the following theorem.

THEOREM 6. The moment-generating function of the normal distribution is given by

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Proof By definition,

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{xt} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[-2xt\sigma^2 + (x-\mu)^2]} dx \end{aligned}$$

and if we complete the square, that is, use the identity

$$-2xt\sigma^2 + (x-\mu)^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

we get

$$M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2} \left\{ \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\frac{x-(\mu+t\sigma^2)}{\sigma}\right]^2} dx \right\}$$

Since the quantity inside the braces is the integral from $-\infty$ to ∞ of a normal density with the parameters $\mu + t\sigma^2$ and σ, and hence is equal to 1, it follows that

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

We are now ready to verify that the parameters μ and σ in Definition 6 are, indeed, the mean and the standard deviation of the normal distribution. Twice differentiating $M_X(t)$ with respect to t, we get

$$M'_X(t) = (\mu + \sigma^2 t) \cdot M_X(t)$$

and

$$M''_X(t) = [(\mu + \sigma^2 t)^2 + \sigma^2] \cdot M_X(t)$$

so that $M'_X(0) = \mu$ and $M''_X(0) = \mu^2 + \sigma^2$. Thus, $E(X) = \mu$ and $\text{var}(X) = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$.

Since the normal distribution plays a basic role in statistics and its density cannot be integrated directly, its areas have been tabulated for the special case where $\mu = 0$ and $\sigma = 1$.

DEFINITION 7. STANDARD NORMAL DISTRIBUTION. *The normal distribution with* $\mu = 0$ *and* $\sigma = 1$ *is referred to as the* ***standard normal distribution****.*

The entries in standard normal distribution table, represented by the shaded area of Figure 5, are the values of

$$\int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

that is, the probabilities that a random variable having the standard normal distribution will take on a value on the interval from 0 to z, for $z = 0.00, 0.01, 0.02, \ldots, 3.08$, and 3.09 and also $z = 4.0$, $z = 5.0$, and $z = 6.0$. By virtue of the symmetry of the normal distribution about its mean, it is unnecessary to extend the table to negative values of z.

EXAMPLE 2

Find the probabilities that a random variable having the standard normal distribution will take on a value

(a) less than 1.72;
(b) less than -0.88;
(c) between 1.30 and 1.75;
(d) between -0.25 and 0.45.

Solution

(a) We look up the entry corresponding to $z = 1.72$ in the standard normal distribution table, add 0.5000 (see Figure 6), and get $0.4573 + 0.5000 = 0.9573$.

(b) We look up the entry corresponding to $z = 0.88$ in the table, subtract it from 0.5000 (see Figure 6), and get $0.5000 - 0.3106 = 0.1894$.

(c) We look up the entries corresponding to $z = 1.75$ and $z = 1.30$ in the table, subtract the second from the first (see Figure 6), and get $0.4599 - 0.4032 = 0.0567$.

(d) We look up the entries corresponding to $z = 0.25$ and $z = 0.45$ in the table, add them (see Figure 6), and get $0.0987 + 0.1736 = 0.2723$.

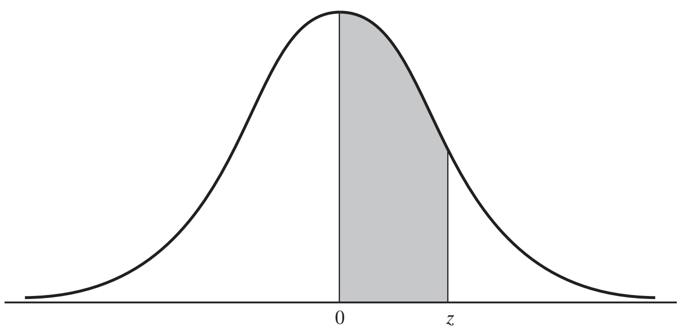

Figure 5. Tabulated areas under the standard normal distribution.

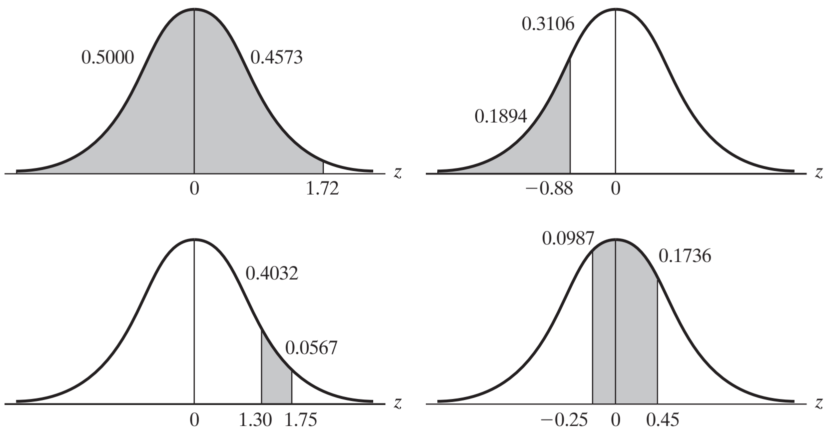

Figure 6. Diagrams for Example 2.

Occasionally, we are required to find a value of z corresponding to a specified probability that falls between values listed in the table. In that case, for convenience, we always choose the z value corresponding to the tabular value that comes closest to the specified probability. However, if the given probability falls midway between tabular values, we shall choose for z the value falling midway between the corresponding values of z.

EXAMPLE 3

With reference to the standard normal distribution table, find the values of z that correspond to entries of

(a) 0.3512;

(b) 0.2533.

Solution

(a) Since 0.3512 falls between 0.3508 and 0.3531, corresponding to $z = 1.04$ and $z = 1.05$, and since 0.3512 is closer to 0.3508 than 0.3531, we choose $z = 1.04$.

(b) Since 0.2533 falls midway between 0.2517 and 0.2549, corresponding to $z = 0.68$ and $z = 0.69$, we choose $z = 0.685$.

To determine probabilities relating to random variables having normal distributions other than the standard normal distribution, we make use of the following theorem.

THEOREM 7. If X has a normal distribution with the mean μ and the standard deviation σ, then

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

has the standard normal distribution.

Proof Since the relationship between the values of X and Z is linear, Z must take on a value between $z_1 = \dfrac{x_1 - \mu}{\sigma}$ and $z_2 = \dfrac{x_2 - \mu}{\sigma}$ when X takes on a value between x_1 and x_2. Hence, we can write

$$\begin{aligned}
P(x_1 < X < x_2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz \\
&= \int_{z_1}^{z_2} n(z; 0, 1)\, dz \\
&= P(z_1 < Z < z_2)
\end{aligned}$$

where Z is seen to be a random variable having the standard normal distribution.

Thus, to use the standard normal distribution table in connection with any random variable having a normal distribution, we simply perform the change of scale $z = \dfrac{x - \mu}{\sigma}$.

EXAMPLE 4

Suppose that the amount of cosmic radiation to which a person is exposed when flying by jet across the United States is a random variable having a normal distribution with a mean of 4.35 mrem and a standard deviation of 0.59 mrem. What is the probability that a person will be exposed to more than 5.20 mrem of cosmic radiation on such a flight?

Solution

Looking up the entry corresponding to $z = \dfrac{5.20 - 4.35}{0.59} = 1.44$ in the table and subtracting it from 0.5000 (see Figure 7), we get $0.5000 - 0.4251 = 0.0749$.

Probabilities relating to random variables having the normal distribution and several other continuous distributions can be found directly with the aid of computer

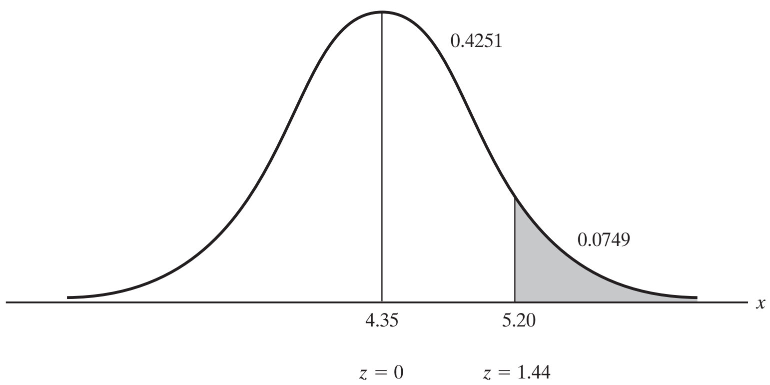

Figure 7. Diagram for Example 4.

programs especially written for statistical applications. The following example illustrates such calculations using MINITAB statistical software.

EXAMPLE 5

Use a computer program to find the probability that a random variable having

(a) the chi-square distribution with 25 degrees of freedom will assume a value greater than 30;

(b) the normal distribution with the mean 18.7 and the standard deviation 9.1 will assume a value between 10.6 and 24.8.

Solution

Using MINITAB software, we select the option "cumulative distribution" to obtain the following:

(a)
```
MTB>CDF C1;
SUBC>Chisquare 25
     3Ø.ØØØØ Ø.7757
```

Thus, the required probability is $1.0000 - 0.7757 = 0.2243$.

(b)
```
MTB>CDF C2;              and   MTB>CDF C3;
SUBC>Normal 18.7 9.1.          SUBC>Normal 18.7 9.1.
     1Ø.6ØØØ Ø.1867                 24.8ØØ Ø.7487
```

Thus, the required probability is $0.7487 - 0.1867 = 0.5620$.

6 The Normal Approximation to the Binomial Distribution

The normal distribution is sometimes introduced as a continuous distribution that provides a close approximation to the binomial distribution when n, the number of trials, is very large and θ, the probability of a success on an individual trial, is close to $\frac{1}{2}$. Figure 8 shows the histograms of binomial distributions with $\theta = \frac{1}{2}$ and $n = 2$,

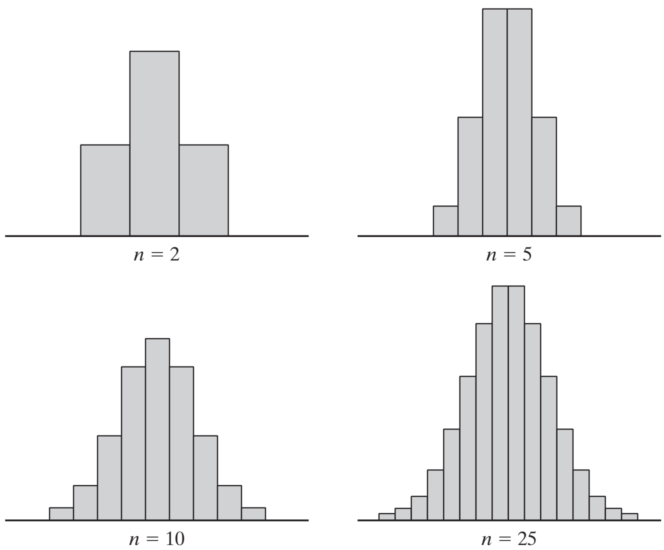

Figure 8. Binomial distributions with $\theta = \frac{1}{2}$.

5, 10, and 25, and it can be seen that with increasing n these distributions approach the symmetrical bell-shaped pattern of the normal distribution.

To provide a theoretical foundation for this argument, let us first prove the following theorem.

THEOREM 8. If X is a random variable having a binomial distribution with the parameters n and θ, then the moment-generating function of

$$Z = \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}$$

approaches that of the standard normal distribution when $n \to \infty$.

Proof Making use of theorems relating to moment-generating functions of the binomial distribution, we can write

$$M_Z(t) = M_{\frac{X-\mu}{\sigma}}(t) = e^{-\mu t/\sigma} \cdot [1 + \theta(e^{t/\sigma} - 1)]^n$$

where $\mu = n\theta$ and $\sigma = \sqrt{n\theta(1-\theta)}$. Then, taking logarithms and substituting the Maclaurin's series of $e^{t/\sigma}$, we get

$$\ln M_{\frac{X-\mu}{\sigma}}(t) = -\frac{\mu t}{\sigma} + n \cdot \ln[1 + \theta(e^{t/\sigma} - 1)]$$

$$= -\frac{\mu t}{\sigma} + n \cdot \ln\left[1 + \theta\left\{\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right\}\right]$$

and, using the infinite series $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$, which converges for $|x| < 1$, to expand this logarithm, it follows that

$$\ln M_{\frac{X-\mu}{\sigma}}(t) = -\frac{\mu t}{\sigma} + n\theta\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right]$$

$$-\frac{n\theta^2}{2}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right]^2$$

$$+\frac{n\theta^3}{3}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right]^3 - \cdots$$

Collecting powers of t, we obtain

$$\ln M_{\frac{X-\mu}{\sigma}}(t) = \left(-\frac{\mu}{\sigma} + \frac{n\theta}{\sigma}\right)t + \left(\frac{n\theta}{2\sigma^2} - \frac{n\theta^2}{2\sigma^2}\right)t^2$$

$$+\left(\frac{n\theta}{6\sigma^3} - \frac{n\theta^2}{2\sigma^3} + \frac{n\theta^3}{3\sigma^3}\right)t^3 + \cdots$$

$$= \frac{1}{\sigma^2}\left(\frac{n\theta - n\theta^2}{2}\right)t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right)t^3 + \cdots$$

since $\mu = n\theta$. Then, substituting $\sigma = \sqrt{n\theta(1-\theta)}$, we find that

$$\ln M_{\frac{X-\mu}{\sigma}}(t) = \frac{1}{2}t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right)t^3 + \cdots$$

For $r > 2$ the coefficient of t^r is a constant times $\frac{n}{\sigma^r}$, which approaches 0 when $n \to \infty$. It follows that

$$\lim_{n\to\infty} \ln M_{\frac{X-\mu}{\sigma}}(t) = \frac{1}{2}t^2$$

and since the limit of a logarithm equals the logarithm of the limit (provided the two limits exist), we conclude that

$$\lim_{n\to\infty} M_{\frac{X-\mu}{\sigma}}(t) = e^{\frac{1}{2}t^2}$$

which is the moment-generating function of Theorem 6 with $\mu = 0$ and $\sigma = 1$.

This completes the proof of Theorem 8, but have we shown that when $n \to \infty$ the distribution of Z, the **standardized** binomial random variable, approaches the standard normal distribution? Not quite. To this end, we must refer to two theorems that we shall state here without proof:

1. *There is a one-to-one correspondence between moment-generating functions and probability distributions (densities) when the former exist.*
2. *If the moment-generating function of one random variable approaches that of another random variable, then the distribution (density) of the first random variable approaches that of the second random variable under the same limiting conditions.*

Strictly speaking, our results apply only when $n \to \infty$, but the normal distribution is often used to approximate binomial probabilities even when n is fairly small. A good rule of thumb is to use this approximation only when $n\theta$ and $n(1-\theta)$ are both greater than 5.

EXAMPLE 6

Use the normal approximation to the binomial distribution to determine the probability of getting 6 heads and 10 tails in 16 flips of a balanced coin.

Solution

To find this approximation, we must use the **continuity correction** according to which each nonnegative integer k is represented by the interval from $k - \frac{1}{2}$ to $k + \frac{1}{2}$. With reference to Figure 9, we must thus determine the area under the curve between 5.5 and 6.5, and since $\mu = 16 \cdot \frac{1}{2} = 8$ and $\sigma = \sqrt{16 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 2$, we must find the area between

$$z = \frac{5.5 - 8}{2} = -1.25 \quad \text{and} \quad z = \frac{6.5 - 8}{2} = -0.75$$

The entries in the standard normal distribution table corresponding to $z = 1.25$ and $z = 0.75$ are 0.3944 and 0.2734, and we find that the normal approximation to the

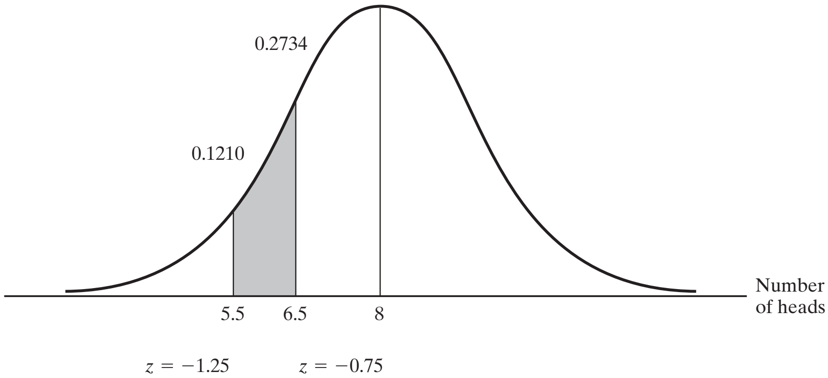

Figure 9. Diagram for Example 6.

probability of "6 heads and 10 tails" is $0.3944 - 0.2734 = 0.1210$. Since the corresponding value in the binomial probabilities table of "Statistical Tables" is 0.1222, we find that the error of the approximation is -0.0012 and that the percentage error is $\frac{0.0012}{0.1222} \cdot 100 = 0.98\%$ in absolute value.

The normal approximation to the binomial distribution used to be applied quite extensively, particularly in approximating probabilities associated with large sets of values of binomial random variables. Nowadays, most of this work is done with computers, as illustrated in Example 5, and we have mentioned the relationship between the binomial and normal distributions primarily because of its theoretical applications.

Exercises

31. Show that the normal distribution has
(a) a relative maximum at $x = \mu$;
(b) inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$.

32. Show that the differential equation of Exercise 30 with $b = c = 0$ and $a > 0$ yields a normal distribution.

33. This question has been intentionally omitted for this edition.

34. If X is a random variable having a normal distribution with the mean μ and the standard deviation σ, find the moment-generating function of $Y = X - c$, where c is a constant, and use it to rework Exercise 33.

35. This question has been intentionally omitted for this edition.

36. This question has been intentionally omitted for this edition.

37. If X is a random variable having the standard normal distribution and $Y = X^2$, show that $\text{cov}(X, Y) = 0$ even though X and Y are evidently not independent.

38. Use the Maclaurin's series expansion of the moment-generating function of the standard normal distribution to show that
(a) $\mu_r = 0$ when r is odd;
(b) $\mu_r = \dfrac{r!}{2^{r/2}\left(\dfrac{r}{2}\right)!}$ when r is even.

39. If we let $K_X(t) = \ln M_{X-\mu}(t)$, the coefficient of $\frac{t^r}{r!}$ in the Maclaurin's series of $K_X(t)$ is called the ***r*th cumulant**, and it is denoted by κ_r. Equating coefficients of like powers, show that
(a) $\kappa_2 = \mu_2$;
(b) $\kappa_3 = \mu_3$;
(c) $\kappa_4 = \mu_4 - 3\mu_2^2$.

40. With reference to Exercise 39, show that for normal distributions $\kappa_2 = \sigma^2$ and all other cumulants are zero.

41. Show that if X is a random variable having the Poisson distribution with the parameter λ and $\lambda \to \infty$, then the moment-generating function of

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

that is, that of a standardized Poisson random variable, approaches the moment-generating function of the standard normal distribution.

42. Show that when $\alpha \to \infty$ and β remains constant, the moment-generating function of a standardized gamma random variable approaches the moment-generating function of the standard normal distribution.

7 The Bivariate Normal Distribution

Among multivariate densities, of special importance is the **multivariate normal distribution**, which is a generalization of the normal distribution in one variable. As it is best (indeed, virtually necessary) to present this distribution in matrix notation, we shall give here only the **bivariate** case; discussions of the general case are listed among the references at the end of this chapter.

DEFINITION 8. BIVARIATE NORMAL DISTRIBUTION. *A pair of random variables* X *and* Y *have a* ***bivariate normal distribution*** *and they are referred to as jointly normally distributed random variables if and only if their joint probability density is given by*

$$f(x,y) = \frac{e^{-\frac{1}{2(1-\rho)^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

for $-\infty < x < \infty$ *and* $-\infty < y < \infty$, *where* $\sigma_1 > 0$, $\sigma_2 > 0$, *and* $-1 < \rho < 1$.

To study this joint distribution, let us first show that the parameters μ_1, μ_2, σ_1, and σ_2 are, respectively, the means and the standard deviations of the two random variables X and Y. To begin with, we integrate on y from $-\infty$ to ∞, getting

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]} dy$$

for the marginal density of X. Then, temporarily making the substitution $u = \frac{x-\mu_1}{\sigma_1}$ to simplify the notation and changing the variable of integration by letting $v = \frac{y-\mu_2}{\sigma_2}$, we obtain

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}\mu^2}}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(v^2 - 2\rho uv)} dv$$

After completing the square by letting

$$v^2 - 2\rho uv = (v - \rho u)^2 - \rho^2 u^2$$

and collecting terms, this becomes

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{v-\rho u}{\sqrt{1-\rho^2}}\right)^2} dv \right\}$$

Finally, identifying the quantity in parentheses as the integral of a normal density from $-\infty$ to ∞, and hence equaling 1, we get

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

for $-\infty < x < \infty$. It follows by inspection that the marginal density of X is a normal distribution with the mean μ_1 and the standard deviation σ_1 and, by symmetry, that the marginal density of Y is a normal distribution with the mean μ_2 and the standard deviation σ_2.

As far as the parameter ρ is concerned, where ρ is the lowercase Greek letter *rho*, it is called the **correlation coefficient**, and the necessary integration will show that $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$. Thus, the parameter ρ measures how the two random variables X and Y vary together.

When we deal with a pair of random variables having a bivariate normal distribution, their conditional densities are also of importance; let us prove the following theorem.

THEOREM 9. If X and Y have a bivariate normal distribution, the conditional density of Y given $X = x$ is a normal distribution with the mean

$$\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

and the variance

$$\sigma^2_{Y|x} = \sigma^2_2(1 - \rho^2)$$

and the conditional density of X given $Y = y$ is a normal distribution with the mean

$$\mu_{X|y} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

and the variance

$$\sigma^2_{X|y} = \sigma^2_1(1 - \rho^2)$$

Proof Writing $w(y|x) = \dfrac{f(x, y)}{g(x)}$ in accordance with the definition of conditional density and letting $u = \dfrac{x - \mu_1}{\sigma_1}$ and $v = \dfrac{y - \mu_2}{\sigma_2}$ to simplify the notation, we get

$$\begin{aligned}
w(y|x) &= \frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2]}}{\dfrac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}u^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[v^2 - 2\rho uv + \rho^2 u^2]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\frac{v - \rho u}{\sqrt{1-\rho^2}}\right]^2}
\end{aligned}$$

Then, expressing this result in terms of the original variables, we obtain

$$w(y|x) = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}}e^{-\frac{1}{2}\left[\frac{y-\left\{\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}}{\sigma_2\sqrt{1-\rho^2}}\right]^2}$$

for $-\infty < y < \infty$, and it can be seen by inspection that this is a normal density with the mean $\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$ and the variance $\sigma^2_{Y|x} = \sigma^2_2(1-\rho^2)$. The corresponding results for the conditional density of X given $Y = y$ follow by symmetry.

The bivariate normal distribution has many important properties, some statistical and some purely mathematical. Among the former, there is the following property, which the reader will be asked to prove in Exercise 43.

THEOREM 10. If two random variables have a bivariate normal distribution, they are independent if and only if $\rho = 0$.

In this connection, if $\rho = 0$, the random variables are said to be **uncorrelated**.

Also, we have shown that for two random variables having a bivariate normal distribution the two marginal densities are normal, but the converse is not necessarily true. In other words, the marginal distributions may both be normal without the joint distribution being a bivariate normal distribution. For instance, if the bivariate density of X and Y is given by

$$f^*(x, y) = \begin{cases} 2f(x, y) & \text{inside squares 2 and 4 of Figure 10} \\ 0 & \text{inside squares 1 and 3 of Figure 10} \\ f(x, y) & \text{elsewhere} \end{cases}$$

where $f(x, y)$ is the value of the bivariate normal density with $\mu_1 = 0, \mu_2 = 0$, and $\rho = 0$ at (x, y), it is easy to see that the marginal densities of X and Y are normal even though their joint density is not a bivariate normal distribution.

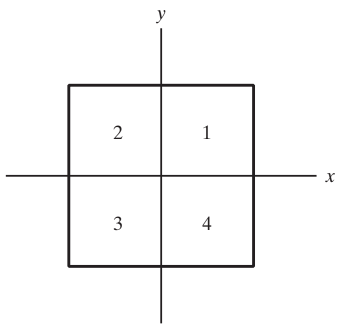

Figure 10. Sample space for the bivariate density given by $f^*(x, y)$.

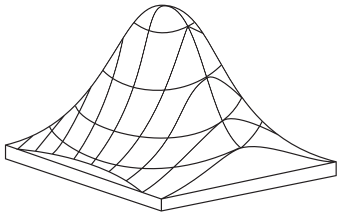

Figure 11. Bivariate normal surface.

Many interesting properties of the bivariate normal density are obtained by studying the **bivariate normal surface**, pictured in Figure 11, whose equation is $z = f(x, y)$, where $f(x, y)$ is the value of the bivariate normal density at (x, y). As the reader will be asked to verify in some of the exercises that follow, the bivariate normal surface has a maximum at (μ_1, μ_2), any plane parallel to the z-axis intersects the surface in a curve having the shape of a normal distribution, and any plane parallel to the xy-plane that intersects the surface intersects it in an ellipse called a **contour of constant probability density**. When $\rho = 0$ and $\sigma_1 = \sigma_2$, the contours of constant probability density are circles, and it is customary to refer to the corresponding joint density as a **circular normal distribution**.

Exercises

43. To prove Theorem 10, show that if X and Y have a bivariate normal distribution, then
(a) their independence implies that $\rho = 0$;
(b) $\rho = 0$ implies that they are independent.

44. Show that any plane perpendicular to the xy-plane intersects the bivariate normal surface in a curve having the shape of a normal distribution.

45. If the exponent of e of a bivariate normal density is

$$\frac{-1}{102}[(x+2)^2 - 2.8(x+2)(y-1) + 4(y-1)^2]$$

find
(a) $\mu_1, \mu_2, \sigma_1, \sigma_2$, and ρ;
(b) $\mu_{Y|x}$ and $\sigma^2_{Y|x}$.

46. If the exponent of e of a bivariate normal density is

$$\frac{-1}{54}(x^2 + 4y^2 + 2xy + 2x + 8y + 4)$$

find σ_1, σ_2, and ρ, given that $\mu_1 = 0$ and $\mu_2 = -1$.

47. If X and Y have the bivariate normal distribution with $\mu_1 = 2, \mu_2 = 5, \sigma_1 = 3, \sigma_2 = 6$, and $\rho = \frac{2}{3}$, find $\mu_{Y|1}$ and $\sigma_{Y|1}$.

48. If X and Y have a bivariate normal distribution and $U = X + Y$ and $V = X - Y$, find an expression for the correlation coefficient of U and V.

49. If X and Y have a bivariate normal distribution, it can be shown that their joint moment-generating function is given by

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E(e^{t_1 X + t_2 Y}) \\ &= e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)} \end{aligned}$$

Verify that
(a) the first partial derivative of this function with respect to t_1 at $t_1 = 0$ and $t_2 = 0$ is μ_1;
(b) the second partial derivative with respect to t_1 at $t_1 = 0$ and $t_2 = 0$ is $\sigma_1^2 + \mu_1^2$;
(c) the second partial derivative with respect to t_1 and t_2 at $t_1 = 0$ and $t_2 = 0$ is $\rho\sigma_1\sigma_2 + \mu_1\mu_2$.

8 The Theory in Practice

In many of the applications of statistics it is assumed that the data are approximately normally distributed. Thus, it is important to make sure that the assumption of normality can, at least reasonably, be supported by the data. Since the normal distribution is symmetric and bell-shaped, examination of the histogram picturing the frequency distribution of the data is useful in checking the assumption of normality. If the histogram is not symmetric, or if it is symmetric but not bell-shaped, the assumption that the data set comes from a normal distribution cannot be supported. Of course, this method is subjective; data that appear to have symmetric, bell-shaped histograms may not be normally distributed.

Another somewhat less subjective method for checking data is the **normal-scores plot**. This plot makes use of ordinary graph paper. It is based on the calculation of **normal scores**, z_p. If n observations are ordered from smallest to largest, they divide the area under the normal curve into $n+1$ equal parts, each having the area $1/(n+1)$. The normal score for the first of these areas is the value of z such that the area under the standard normal curve to the left of z is $1/(n+1)$, or $-z_{1/(n+1)}$. Thus, the normal scores for $n = 4$ observations are $-z_{0.20} = -0.84, -z_{0.40} = -0.25, z_{0.40} = 0.25$, and $z_{20} = 0.84$. The ordered observations then are plotted against the corresponding normal scores on ordinary graph paper.

EXAMPLE 7

Find the normal scores and the coordinates for making a normal-scores plot of the following six observations:

$$3, \quad 2, \quad 7, \quad 4, \quad 3, \quad 5$$

Solution

Since $n = 6$, there are 6 normal scores, as follows: $-z_{0.14} = -1.08, -z_{0.29} = -0.55, -z_{0.43} = -0.18, z_{0.43} = 0.18, z_{0.29} = 0.55$, and $z_{0.14} = 1.08$. When the observations are ordered and tabulated together with the normal scores, the following table results:

Observation:	2	3	3	4	5	7
Normal score:	−1.08	−0.55	−0.18	0.18	0.55	1.08

The coordinates for a normal-scores plot make use of a cumulative percentage distribution of the data. The cumulative percentage distribution is as follows:

Class Boundary	*Cumulative Percentage*	*Normal Score*
4395	5	−1.64
4595	17	−0.95
4795	37	−0.33
4995	69	0.50
5195	87	1.13
5395	97	1.88

A graph of the class boundaries versus the normal scores is shown in Figure 12. It can be seen from this graph that the points lie in an almost perfect straight line, strongly suggesting that the underlying data are very close to being normally distributed.

In modern practice, use of MINITAB or other statistical software eases the computation considerably. In addition, MINITAB offers three tests for normality that are less subjective than mere examination of a normal-scores plot.

Sometimes a normal-scores plot showing a curve can be changed to a straight line by means of an appropriate transformation. The procedure involves identifying

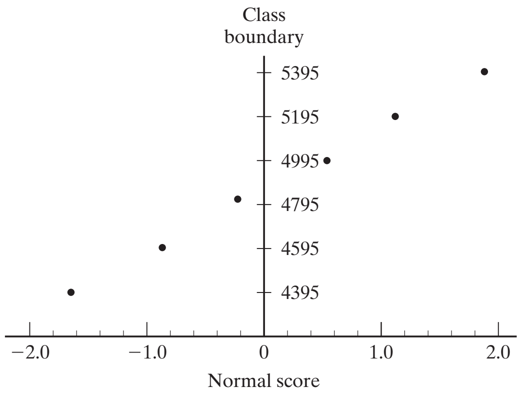

Figure 12. Normal-scores plot.

the type of transformation needed, making the transformation, and then checking the transformed data by means of a normal-scores plot to see if they can be assumed to have a normal distribution.

When data appear not to be normally distributed because of *too many large values*, the following transformations are good candidates to try:

$$\text{logarithmic transformation } u = \log(x)$$
$$\text{square-root transformation } u = \sqrt{x}$$
$$\text{reciprocal transformation } u = \frac{1}{x}$$

When data exhibit *too many small values*, the following transformations may produce approximately normal data:

$$\text{power transformation} \qquad u = x^a, \text{ where } a > 1$$
$$\text{exponential transformation} \qquad u = a^x, \text{ where } a > 1$$

On rare occasions, it helps to make a linear transformation of the form $u = a + bx$ first, and then to use one of the indicated transformations. This strategy becomes necessary when some of the data have negative values and logarithmic, square-root, or certain power transformations are to be tried. However, making a linear transformation alone cannot be effective. If x is a value of a normally distributed random variable, then the random variable having the values $a + bx$ also has the normal distribution. Thus, a linear transformation alone cannot transform nonnormally distributed data into normality.

EXAMPLE 8

Make a normal-scores plot of the following data. If the plot does not appear to show normality, make an appropriate transformation, and check the transformed data for normality.

54.9 8.3 5.2 32.4 15.5

Solution
The normal scores are $-0.95, -0.44, 0, 0.44$, and 0.95. A normal-scores plot of these data (Figure 13[a]) shows sharp curvature. Since two of the five values are very large compared with the other three values, a logarithmic transformation (base 10) was used to transform the data to

$$1.74 \quad 0.92 \quad 0.72 \quad 1.51 \quad 1.19$$

A normal-scores plot of these transformed data (Figure 13[b]) shows a nearly straight line, indicating that the transformed data are approximately normally distributed.

If lack of normality seems to result from one or a small number of aberrant observations called outliers, a single large observation, a single small observation, or both, it is not likely that the data can be transformed to normality. It is difficult to give a hard-and-fast rule for identifying outliers. For example, it may be inappropriate to define an **outlier** as an observation whose value is more than three standard deviations from the mean, since such an observation can occur with a reasonable probability in a large number of observations taken from a normal distribution. Ordinarily, an observation that clearly does not lie on the straight line defined by the other observations in a normal-scores plot can be considered an outlier. In the presence of suspected outliers, it is customary to examine normal-scores plots of the data after the outlier or outliers have been omitted.

Outlying observations may result from several causes, such as an error in recording data, an error of observation, or an unusual event such as a particle of dust settling on a material during thin-film deposition. There is always a great temptation to drop outliers from a data set entirely on the basis that they do not seem to belong to the main body of data. But an outlier can be as informative about the process from which the data were taken as the remainder of the data. Outliers which occur infrequently, but regularly in successive data sets, give evidence that should not be ignored. For example, a hole with an unusually large diameter might result from a drill not having been inserted properly into the chuck. Perhaps the condition was corrected after one or two holes were drilled, and the operator failed to discard the parts with the "bad" hole, thus producing one or two outliers. While outliers sometimes are separated from the other data for the purpose of performing a preliminary

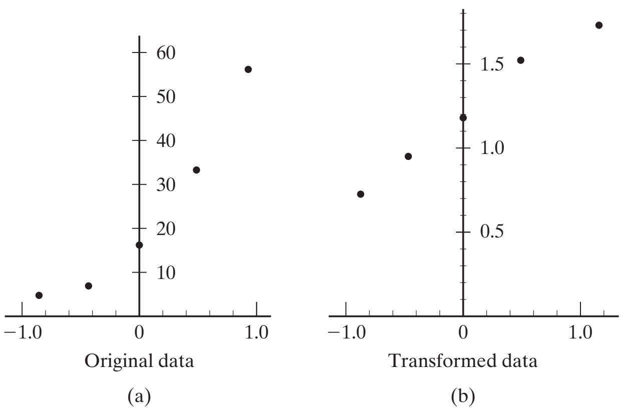

Figure 13. Normal-scores plot for Example 8.

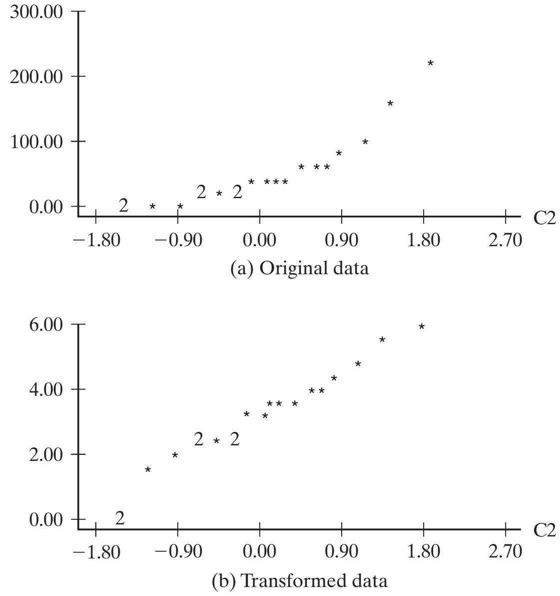

Figure 14. Normal-scores plots.

analysis, they should be discarded only after a good reason for their existence has been found.

Normal scores and normal-score plots can be obtained with a variety of statistical software. To illustrate the procedure using MINITAB, 20 numbers are entered with the following command and data-entry instructions

```
SET C1:
0 215 31 7 15 80 17 41 51 3 58 158 0 11 42 11 17 32 64 100
END
```

Then the command `NSCORES C1 PUT IN C2` is given to find the normal scores and place them in the second column. A normal-scores plot, generated by the command `PLOT C1 VS C2`, is shown in Figure 14(a). The points in this graph clearly do not follow a straight line. Several power transformations were tried in an attempt to transform the data to normality. The cube-root transformation $u = X^{1/3}$, made by giving the command `RAISE C1 TO THE POWER .3333 PUT IN C3`, seemed to work best. Then, a normal-scores plot of the transformed data was generated with the command `PLOT C3 VS C2`, as shown in Figure 14(b). It appears from this graph that the cube roots of the original data are approximately normally distributed.

Applied Exercises — SECS. 1–4

50. In certain experiments, the error made in determining the density of a substance is a random variable having a uniform density with $\alpha = -0.015$ and $\beta = 0.015$. Find the probabilities that such an error will
(a) be between -0.002 and 0.003;
(b) exceed 0.005 in absolute value.

51. A point D is chosen on the line AB, whose midpoint is C and whose length is a. If X, the distance from D to A, is a random variable having the uniform density with $\alpha = 0$ and $\beta = a$, what is the probability that AD, BD, and AC will form a triangle?

52. In a certain city, the daily consumption of electric power in millions of kilowatt-hours can be treated as a random variable having a gamma distribution with $\alpha = 3$ and $\beta = 2$. If the power plant of this city has a daily capacity of 12 million kilowatt-hours, what is the probability that this power supply will be inadequate on any given day?

53. If a company employs n salespersons, its gross sales in thousands of dollars may be regarded as a random variable having a gamma distribution with $\alpha = 80\sqrt{n}$ and $\beta = 2$. If the sales cost is \$8,000 per salesperson, how many salespersons should the company employ to maximize the expected profit?

54. The amount of time that a watch will run without having to be reset is a random variable having an exponential distribution with $\theta = 120$ days. Find the probabilities that such a watch will
(a) have to be reset in less than 24 days;
(b) not have to be reset in at least 180 days.

55. The mileage (in thousands of miles) that car owners get with a certain kind of radial tire is a random variable having an exponential distribution with $\theta = 40$. Find the probabilities that one of these tires will last
(a) at least 20,000 miles;
(b) at most 30,000 miles.

56. The number of bad checks that a bank receives during a 5-hour business day is a Poisson random variable with $\lambda = 2$. What is the probability that it will not receive a bad check on any one day during the first 2 hours of business?

57. The number of planes arriving per day at a small private airport is a random variable having a Poisson distribution with $\lambda = 28.8$. What is the probability that the time between two such arrivals is at least 1 hour?

58. If the annual proportion of erroneous income tax returns filed with the IRS can be looked upon as a random variable having a beta distribution with $\alpha = 2$ and $\beta = 9$, what is the probability that in any given year there will be fewer than 10 percent erroneous returns?

59. A certain kind of appliance requires repairs on the average once every 2 years. Assuming that the times between repairs are exponentially distributed, what is the probability that such an appliance will work at least 3 years without requiring repairs?

60. If the annual proportion of new restaurants that fail in a given city may be looked upon as a random variable having a beta distribution with $\alpha = 1$ and $\beta = 4$, find
(a) the mean of this distribution, that is, the annual proportion of new restaurants that can be expected to fail in the given city;
(b) the probability that at least 25 percent of all new restaurants will fail in the given city in any one year.

61. Suppose that the service life in hours of a semiconductor is a random variable having a Weibull distribution (see Exercise 23) with $\alpha = 0.025$ and $\beta = 0.500$.
(a) How long can such a semiconductor be expected to last?
(b) What is the probability that such a semiconductor will still be in operating condition after 4,000 hours?

SECS. 5–7

62. If Z is a random variable having the standard normal distribution, find
(a) $P(Z < 1.33)$;
(b) $P(Z \geqq -0.79)$;
(c) $P(0.55 < Z < 1.22)$;
(d) $P(-1.90 \leq Z \leq 0.44)$.

63. If Z is a random variable having the standard normal distribution, find the probabilities that it will take on a value
(a) greater than 1.14;
(b) greater than -0.36;
(c) between -0.46 and -0.09;
(d) between -0.58 and 1.12.

64. If Z is a random variable having the standard normal distribution, find the respective values z_1, z_2, z_3, and z_4 such that
(a) $P(0 < Z < z_1) = 0.4306$;
(b) $P(Z \geqq z_2) = 0.7704$;
(c) $P(Z > z_3) = 0.2912$;
(d) $P(-z_4 \leqq Z < z_4) = 0.9700$.

65. Find z if the standard-normal-curve area
(a) between 0 and z is 0.4726;
(b) to the left of z is 0.9868;
(c) to the right of z is 0.1314;
(d) between $-z$ and z is 0.8502.

66. If X is a random variable having a normal distribution, what are the probabilities of getting a value
(a) within one standard deviation of the mean;
(b) within two standard deviations of the mean;
(c) within three standard deviations of the mean;
(d) within four standard deviations of the mean?

67. If z_α is defined by

$$\int_{z_\alpha}^{\infty} n(z; 0, 1)\, dz = \alpha$$

find its values for
(a) $\alpha = 0.05$;
(b) $\alpha = 0.025$;
(c) $\alpha = 0.01$;
(d) $\alpha = 0.005$.

68. (a) Use a computer program to find the probability that a random variable having the normal distribution with the mean -1.786 and the standard deviation 1.0416 will assume a value between -2.159 and 0.5670.

(b) Interpolate in the standard normal distribution table to find this probability and compare your result with the more exact value found in part (a).

69. (a) Use a computer program to find the probability that a random variable having the normal distribution with mean 5.853 and the standard deviation 1.361 will assume a value greater than 8.625.

(b) Interpolate in the standard normal distribution table to find this probability and compare your result with the more exact value found in part (a).

70. Suppose that during periods of meditation the reduction of a person's oxygen consumption is a random variable having a normal distribution with $\mu = 37.6$ cc per minute and $\sigma = 4.6$ cc per minute. Find the probabilities that during a period of meditation a person's oxygen consumption will be reduced by

(a) at least 44.5 cc per minute;

(b) at most 35.0 cc per minute;

(c) anywhere from 30.0 to 40.0 cc per minute.

71. In a photographic process, the developing time of prints may be looked upon as a random variable having the normal distribution with $\mu = 15.40$ seconds and $\sigma = 0.48$ second. Find the probabilities that the time it takes to develop one of the prints will be

(a) at least 16.00 seconds;

(b) at most 14.20 seconds;

(c) anywhere from 15.00 to 15.80 seconds.

72. A random variable has a normal distribution with $\sigma = 10$. If the probability that the random variable will take on a value less than 82.5 is 0.8212, what is the probability that it will take on a value greater than 58.3?

73. Suppose that the actual amount of instant coffee that a filling machine puts into "6-ounce" jars is a random variable having a normal distribution with $\sigma = 0.05$ ounce. If only 3 percent of the jars are to contain less than 6 ounces of coffee, what must be the mean fill of these jars?

74. Check in each case whether the normal approximation to the binomial distribution may be used according to the rule of thumb in Section 6.

(a) $n = 16$ and $\theta = 0.20$;

(b) $n = 65$ and $\theta = 0.10$;

(c) $n = 120$ and $\theta = 0.98$.

75. Suppose that we want to use the normal approximation to the binomial distribution to determine $b(1; 150, 0.05)$.

(a) Based on the rule of thumb in Section 6, would we be justified in using the approximation?

(b) Make the approximation and round to four decimals.

(c) If a computer printout shows that $b(1; 150, 0.05) = 0.0036$ rounded to four decimals, what is the percentage error of the approximation obtained in part (b)?

This serves to illustrate that the rule of thumb is just that and no more; making approximations like this also requires a good deal of professional judgment.

76. Use the normal approximation to the binomial distribution to determine (to four decimals) the probability of getting 7 heads and 7 tails in 14 flips of a balanced coin. Also refer to the binomial probabilities table of "Statistical Tables" to find the error of this approximation.

77. With reference to Exercise 75, show that the Poisson distribution would have yielded a better approximation.

78. If 23 percent of all patients with high blood pressure have bad side effects from a certain kind of medicine, use the normal approximation to find the probability that among 120 patients with high blood pressure treated with this medicine more than 32 will have bad side effects.

79. If the probability is 0.20 that a certain bank will refuse a loan application, use the normal approximation to determine (to three decimals) the probability that the bank will refuse at most 40 of 225 loan applications.

80. To illustrate the law of large numbers, use the normal approximation to the binomial distribution to determine the probabilities that the proportion of heads will be anywhere from 0.49 to 0.51 when a balanced coin is flipped

(a) 100 times;

(b) 1,000 times;

(c) 10,000 times.

SEC. 8

81. Check the following data for normality by finding normal scores and making a normal-scores plot:

3.9 4.6 4.5 1.6 4.2

82. Check the following data for normality by finding normal scores and making a normal-scores plot:

36 22 3 13 31 45

83. This question has been intentionally omitted for this edition.

84. The weights (in pounds) of seven shipments of bolts are

37 45 11 51 13 48 61

Make a normal-scores plot of these weights. Can they be regarded as having come from a normal distribution?

85. This question has been intentionally omitted for this edition.

86. Use a computer program to make a normal-scores plot for the data on the time to make coke in successive runs of a coke oven (given in hours).

7.8	9.2	6.4	8.2	7.6	5.9	7.4	7.1	6.7	8.5
10.1	8.6	7.7	5.9	9.3	6.4	6.8	7.9	7.2	10.2
6.9	7.4	7.8	6.6	8.1	9.5	6.4	7.6	8.4	9.2

Also test these data for normality using the three tests given by MINITAB.

87. Eighty pilots were tested in a flight simulator and the time for each to take corrective action for a given emergency was measured in seconds, with the following results:

11.1	5.2	3.6	7.6	12.4	6.8	3.8	5.7	9.0	6.0	4.9	12.6
7.4	5.3	14.2	8.0	12.6	13.7	3.8	10.6	6.8	5.4	9.7	6.7
14.1	5.3	11.1	13.4	7.0	8.9	6.2	8.3	7.7	4.5	7.6	5.0
9.4	3.5	7.9	11.0	8.6	10.5	5.7	7.0	5.6	9.1	5.1	4.5
6.2	6.8	4.3	8.5	3.6	6.1	5.8	10.0	6.4	4.0	5.4	7.0
4.1	8.1	5.8	11.8	6.1	9.1	3.3	12.5	8.5	10.8	6.5	7.9
6.8	10.1	4.9	5.4	9.6	8.2	4.2	3.4				

Use a computer to make a normal-scores plot of these data, and test for normality.

References

Useful information about various special probability densities, in outline form, may be found in

Derman, C., Gleser, L., and Olkin, I., *Probability Models and Applications*. New York: Macmillan Publishing Co., Inc., 1980,

Hastings, N. A. J., and Peacock, J. B., *Statistical Distributions*. London: Butterworth and Co. Ltd., 1975,

and

Johnson, N. L., and Kotz, S., *Continuous Univariate Distributions*, Vols. 1 and 2. Boston: Houghton Mifflin Company, 1970.

A direct proof that the standardized binomial distribution approaches the standard normal distribution when $n \to \infty$ is given in

Keeping, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962.

A detailed treatment of the mathematical and statistical properties of the bivariate normal surface may be found in

Yule, G. U., and Kendall, M. G., *An Introduction to the Theory of Statistics*, 14th ed. New York: Hafner Publishing Co., Inc., 1950.

The multivariate normal distribution is treated in matrix notation in

Bickel, P. J., and Doksum, K. A., *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco: Holden-Day, Inc., 1977,

Hogg, R. V., and Craig, A. T., *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan Publishing Co., Inc., 1978,

Lindgren, B. W., *Statistical Theory*, 3rd ed. New York: Macmillan Publishing Co., Inc., 1976.

Answers to Odd-Numbered Exercises

3 $F(x) = \begin{cases} 0 & \text{for } x \le \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \text{for } \alpha < x < \beta \\ 1 & \text{for } x \ge \beta \end{cases}$

5 $\alpha_3 = 0$ and $\alpha_4 = \frac{9}{5}$.

11 For $0 < \alpha < 1$ the function $\to \infty$ when $x \to 0$; for $\alpha = 1$ the function has an absolute maximum at $x = 0$.

13 $\mu_1' = \alpha\beta, \mu_2' = \alpha(\alpha+1)\beta^2, \mu_3' = \alpha(\alpha+1)(\alpha+2)\beta^3$, and $\mu_4' = \alpha(\alpha+1)(\alpha+2)(\alpha+3)\beta^4$.

17 $M_Y(t) = \dfrac{e^{-\theta t}}{1-\theta t}$.

19 For $0 < \nu < 2$ the function $\to \infty$ when $x \to 0$, for $\nu = 2$ the function has an absolute maximum at $x = 0$.

23 **(a)** $k = \alpha\beta$.

33 $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$.

45 **(a)** $\mu_1 = -2$, $\mu_2 = 1$, $\sigma_1 = 10$, $\sigma_2 = 5$, and $\rho = 0.7$.

47 $\mu_{Y|1} = \frac{11}{3}$, $\sigma_{Y|1} = \sqrt{20} = 4.47$.

51 $\frac{1}{2}$.

53 n = 100.

55 **(a)** 0.6065; **(b)** 0.5276.

57 0.1827.

59 0.2231.

61 **(a)** 3200 hours; **(b)** 0.2060.

63 **(a)** 0.1271; **(b)** 0.6406; **(c)** 0.1413; **(d)** 0.5876.

65 **(a)** 1.92; **(b)** 2.22; **(c)** 1.12; **(d)** ±1.44.

67 **(a)** 1.645; **(b)** 1.96; **(c)** 2.33; **(d)** 2.575.

69 **(a)** 0.0208.

71 **(a)** 0.1056; **(b)** 0.0062; **(c)** 0.5934.

73 6.094 ounces.

75 **(a)** yes; **(b)** 0.0078; **(c)** 117%.

77 0.0041.

79 0.227.

This page intentionally left blank

Functions of Random Variables

1 Introduction

In this chapter we shall concern ourselves with the problem of finding the probability distributions or densities of **functions of one or more random variables**. That is, given a set of random variables $X_1, X_2, \ldots, X_n$ and their joint probability distribution or density, we shall be interested in finding the probability distribution or density of some random variable $Y = u(X_1, X_2, \ldots, X_n)$. This means that the values of Y are related to those of the X's by means of the equation

$$y = u(x_1, x_2, \ldots, x_n)$$

Several methods are available for solving this kind of problem. The ones we shall discuss in the next four sections are called the **distribution function technique**, the **transformation technique**, and the **moment-generating function technique**. Although all three methods can be used in some situations, in most problems one technique will be preferable (easier to use than the others). This is true, for example, in some instances where the function in question is linear in the random variables $X_1, X_2, \ldots,$ X_n, and the moment-generating function technique yields the simplest derivations.

2 Distribution Function Technique

A straightforward method of obtaining the probability density of a function of continuous random variables consists of first finding its distribution function and then its probability density by differentiation. Thus, if $X_1, X_2, \ldots, X_n$ are continuous random variables with a given joint probability density, the probability density of $Y = u(X_1, X_2, \ldots, X_n)$ is obtained by first determining an expression for the probability

$$F(y) = P(Y \leqq y) = P[u(X_1, X_2, \ldots, X_n) \leqq y]$$

and then differentiating to get

$$f(y) = \frac{dF(y)}{dy}$$

From Chapter 7 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

EXAMPLE 1

If the probability density of X is given by

$$f(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability density of $Y = X^3$.

Solution

Letting $G(y)$ denote the value of the distribution function of Y at y, we can write

$$\begin{aligned} G(y) &= P(Y \leqq y) \\ &= P(X^3 \leqq y) \\ &= P(X \leqq y^{1/3}) \\ &= \int_0^{y^{1/3}} 6x(1-x)\,dx \\ &= 3y^{2/3} - 2y \end{aligned}$$

and hence

$$g(y) = 2(y^{-1/3} - 1)$$

for $0 < y < 1$; elsewhere, $g(y) = 0$. In Exercise 15 the reader will be asked to verify this result by a different technique.

EXAMPLE 2

If $Y = |X|$, show that

$$g(y) = \begin{cases} f(y) + f(-y) & \text{for } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $f(x)$ is the value of the probability density of X at x and $g(y)$ is the value of the probability density of Y at y. Also, use this result to find the probability density of $Y = |X|$ when X has the standard normal distribution.

Solution

For $y > 0$ we have

$$\begin{aligned} G(y) &= P(Y \leqq y) \\ &= P(|X| \leqq y) \\ &= P(-y \leqq X \leqq y) \\ &= F(y) - F(-y) \end{aligned}$$

and, upon differentiation,

$$g(y) = f(y) + f(-y)$$

Also, since $|x|$ cannot be negative, $g(y) = 0$ for $y < 0$. Arbitrarily letting $g(0) = 0$, we can thus write

$$g(y) = \begin{cases} f(y) + f(-y) & \text{for } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

If X has the standard normal distribution and $Y = |X|$, it follows that

$$\begin{aligned} g(y) &= n(y; 0, 1) + n(-y; 0, 1) \\ &= 2n(y; 0, 1) \end{aligned}$$

for $y > 0$ and $g(y) = 0$ elsewhere. An important application of this result may be found in Example 9.

EXAMPLE 3

If the joint density of X_1 and X_2 is given by

$$f(x_1, x_2) = \begin{cases} 6e^{-3x_1 - 2x_2} & \text{for } x_1 > 0, x_2 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability density of $Y = X_1 + X_2$.

Solution

Integrating the joint density over the shaded region of Figure 1, we get

$$\begin{aligned} F(y) &= \int_0^y \int_0^{y - x_2} 6e^{-3x_1 - 2x_2} dx_1\, dx_2 \\ &= 1 + 2e^{-3y} - 3e^{-2y} \end{aligned}$$

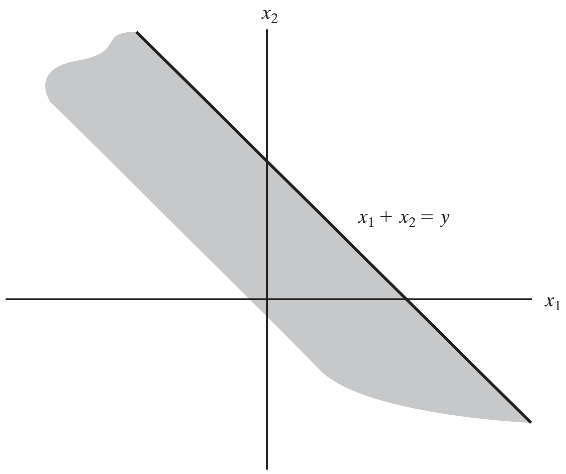

Figure 1. Diagram for Example 3.

and, differentiating with respect to y, we obtain

$$f(y) = 6(e^{-2y} - e^{-3y})$$

for $y > 0$; elsewhere, $f(y) = 0$.

Exercises

1. If X has an exponential distribution with the parameter θ, use the distribution function technique to find the probability density of the random variable $Y = \ln X$.

2. If the probability density of X is given by

$$f(x) = \begin{cases} 2xe^{-x^2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and $Y = X^2$, find
(a) the distribution function of Y;
(b) the probability density of Y.

3. If X has the uniform density with the parameters $\alpha = 0$ and $\beta = 1$, use the distribution function technique to find the probability density of the random variable $Y = \sqrt{X}$.

4. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} 4xye^{-(x^2+y^2)} & \text{for } x > 0, y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and $Z = \sqrt{X^2 + Y^2}$, find
(a) the distribution function of Z;
(b) the probability density of Z.

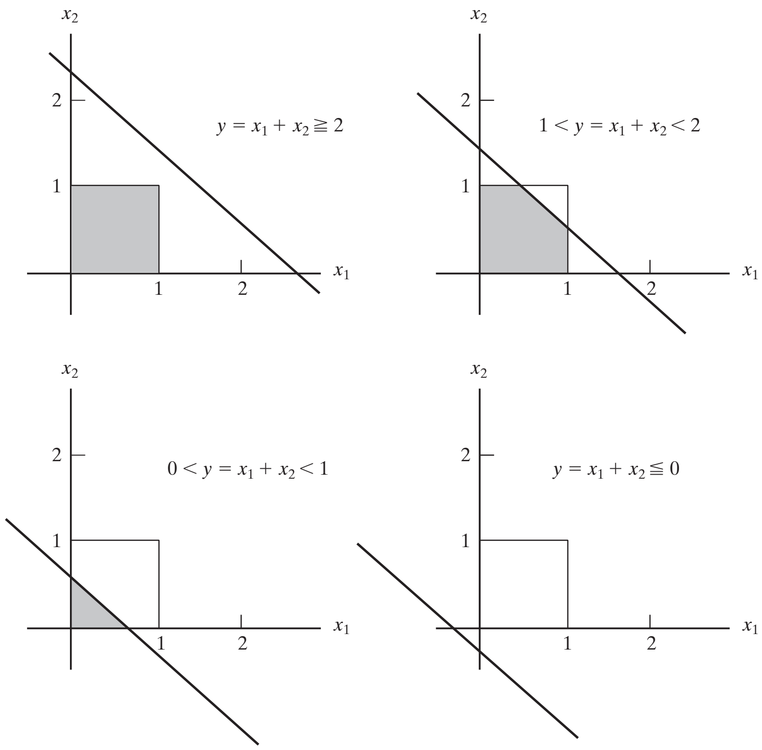

Figure 2. Diagram for Exercise 6.

5. If X_1 and X_2 are independent random variables having exponential densities with the parameters θ_1 and θ_2, use the distribution function technique to find the probability density of $Y = X_1 + X_2$ when

(a) $\theta_1 \neq \theta_2$;
(b) $\theta_1 = \theta_2$.

(Example 3 is a special case of this with $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{1}{2}$.)

6. Let X_1 and X_2 be independent random variables having the uniform density with $\alpha = 0$ and $\beta = 1$. Referring to Figure 2, find expressions for the distribution function of $Y = X_1 + X_2$ for

(a) $y \leqq 0$;
(b) $0 < y < 1$;
(c) $1 < y < 2$;
(d) $y \geqq 2$.

Also find the probability density of Y.

7. With reference to the two random variables of Exercise 5, show that if $\theta_1 = \theta_2 = 1$, the random variable

$$Z = \frac{X_1}{X_1 + X_2}$$

has the uniform density with $\alpha = 0$ and $\beta = 1$.

8. If the joint density of X and Y is given by

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{for } x > 0, y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and $Z = \dfrac{X+Y}{2}$, find the probability density of Z by the distribution function technique.

3 Transformation Technique: One Variable

Let us show how the probability distribution or density of a function of a random variable can be determined without first getting its distribution function. In the discrete case there is no real problem as long as the relationship between the values of X and $Y = u(X)$ is one-to-one; all we have to do is make the appropriate substitution.

EXAMPLE 4

If X is the number of heads obtained in four tosses of a balanced coin, find the probability distribution of $Y = \dfrac{1}{1+X}$.

Solution

Using the formula for the binomial distribution with $n = 4$ and $\theta = \frac{1}{2}$, we find that the probability distribution of X is given by

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Then, using the relationship $y = \dfrac{1}{1+x}$ to substitute values of Y for values of X, we find that the probability distribution of Y is given by

y	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$
$g(y)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

If we had wanted to make the substitution directly in the formula for the binomial distribution with $n = 4$ and $\theta = \frac{1}{2}$, we could have substituted $x = \dfrac{1}{y} - 1$ for x in

$$f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^4 \quad \text{for } x = 0, 1, 2, 3, 4$$

getting

$$g(y) = f\left(\frac{1}{y} - 1\right) = \binom{4}{\frac{1}{y} - 1}\left(\frac{1}{2}\right)^4 \quad \text{for } y = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$$

Note that in the preceding example the probabilities remained unchanged; the only difference is that in the result they are associated with the various values of Y instead of the corresponding values of X. That is all there is to the **transformation** (or **change-of-variable**) **technique** in the discrete case as long as the relationship is one-to-one. If the relationship is not one-to-one, we may proceed as in the following example.

EXAMPLE 5

With reference to Example 4, find the probability distribution of the random variable $Z = (X - 2)^2$.

Solution
Calculating the probabilities $h(z)$ associated with the various values of Z, we get

$$h(0) = f(2) = \frac{6}{16}$$

$$h(1) = f(1) + f(3) = \frac{4}{16} + \frac{4}{16} = \frac{8}{16}$$

$$h(4) = f(0) + f(4) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16}$$

and hence

z	0	1	4
$h(z)$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{1}{8}$

To perform a transformation of variable in the continuous case, we shall assume that the function given by $y = u(x)$ is differentiable and either increasing or decreasing for all values within the range of X for which $f(x) \neq 0$, so the inverse function, given by $x = w(y)$, exists for all the corresponding values of y and is differentiable except where $u'(x) = 0$.† Under these conditions, we can prove the following theorem.

†To avoid points where $u'(x)$ might be 0, we generally do not include the endpoints of the intervals for which probability densities are nonzero. This is the practice that we follow throughout this chapter.

THEOREM 1. Let $f(x)$ be the value of the probability density of the continuous random variable X at x. If the function given by $y = u(x)$ is differentiable and either increasing or decreasing for all values within the range of X for which $f(x) \neq 0$, then, for these values of x, the equation $y = u(x)$ can be uniquely solved for x to give $x = w(y)$, and for the corresponding values of y the probability density of $Y = u(X)$ is given by

$$g(y) = f[w(y)] \cdot |w'(y)| \quad \text{provided } u'(x) \neq 0$$

Elsewhere, $g(y) = 0$.

Proof First, let us prove the case where the function given by $y = u(x)$ is increasing. As can be seen from Figure 3, X must take on a value between $w(a)$ and $w(b)$ when Y takes on a value between a and b. Hence,

$$\begin{aligned} P(a < Y < b) &= P[w(a) < X < w(b)] \\ &= \int_{w(a)}^{w(b)} f(x)\,dx \\ &= \int_a^b f[w(y)]w'(y)\,dy \end{aligned}$$

where we performed the change of variable $y = u(x)$, or equivalently $x = w(y)$, in the integral. The integrand gives the probability density of Y as long as $w'(y)$ exists, and we can write

$$g(y) = f[w(y)]w'(y)$$

When the function given by $y = u(x)$ is decreasing, it can be seen from Figure 3 that X must take on a value between $w(b)$ and $w(a)$ when Y takes on a value between a and b. Hence,

$$\begin{aligned} P(a < Y < b) &= P[w(b) < X < w(a)] \\ &= \int_{w(b)}^{w(a)} f(x)\,dx \\ &= \int_b^a f[w(y)]w'(y)\,dy \\ &= -\int_a^b f[w(y)]w'(y)\,dy \end{aligned}$$

where we performed the same change of variable as before, and it follows that

$$g(y) = -f[w(y)]w'(y)$$

Since $w'(y) = \dfrac{dx}{dy} = \dfrac{1}{\frac{dy}{dx}}$ is positive when the function given by $y = u(x)$ is increasing, and $-w'(y)$ is positive when the function given by $y = u(x)$ is decreasing, we can combine the two cases by writing

$$g(y) = f[w(y)] \cdot |w'(y)|$$

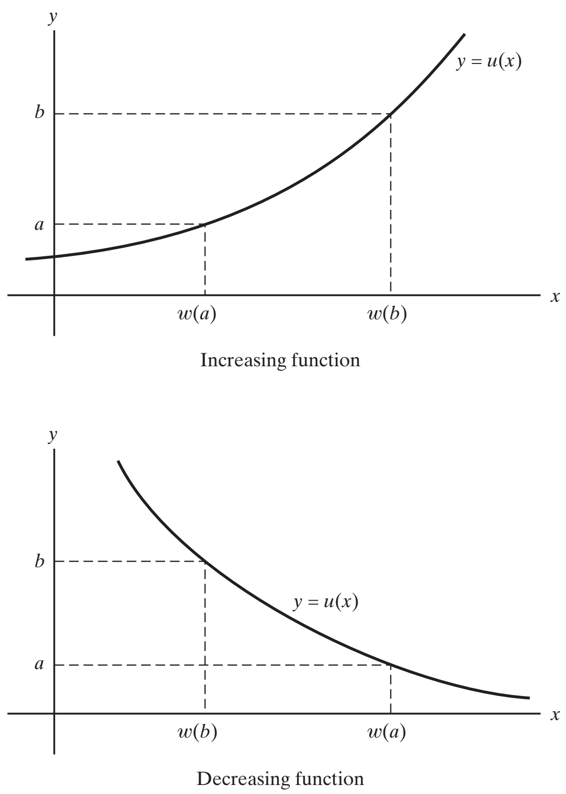

Figure 3. Diagrams for proof of Theorem 1.

EXAMPLE 6

If X has the exponential distribution given by

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability density of the random variable $Y = \sqrt{X}$.

Solution

The equation $y = \sqrt{x}$, relating the values of X and Y, has the unique inverse $x = y^2$, which yields $w'(y) = \dfrac{dx}{dy} = 2y$. Therefore,

$$g(y) = e^{-y^2}|2y| = 2ye^{-y^2}$$

for $y > 0$ in accordance with Theorem 1. Since the probability of getting a value of Y less than or equal to 0, like the probability of getting a value of X less than or equal to 0, is zero, it follows that the probability density of Y is given by

$$g(y) = \begin{cases} 2ye^{-y^2} & \text{for } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

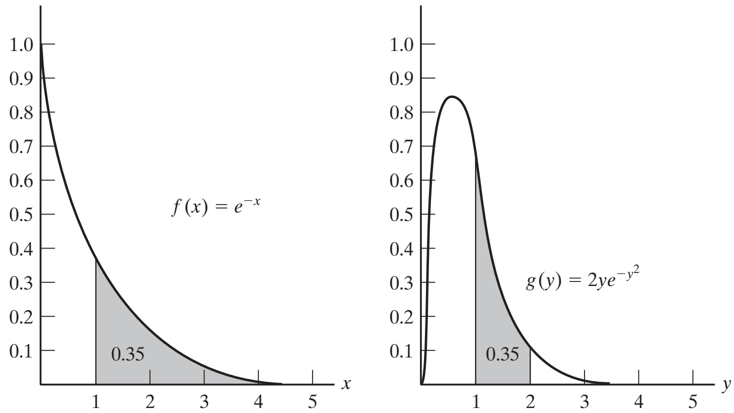

Figure 4. Diagrams for Example 6.

The two diagrams of Figure 4 illustrate what happened in this example when we transformed from X to Y. As in the discrete case (for instance, Example 4), the probabilities remain the same, but they pertain to different values (intervals of values) of the respective random variables. In the diagram on the left, the 0.35 probability pertains to the event that X will take on a value on the interval from 1 to 4, and in the diagram on the right, the 0.35 probability pertains to the event that Y will take on a value on the interval from 1 to 2.

EXAMPLE 7

If the double arrow of Figure 5 is spun so that the random variable Θ has the uniform density

$$f(\theta) = \begin{cases} \frac{1}{\pi} & \text{for } -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{elsewhere} \end{cases}$$

determine the probability density of X, the abscissa of the point on the x-axis to which the arrow will point.

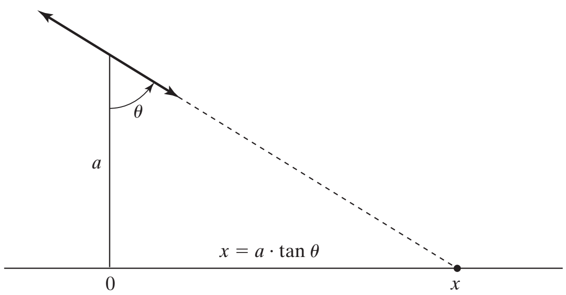

Figure 5. Diagram for Example 7.

Solution
As is apparent from the diagram, the relationship between x and θ is given by $x = a \cdot \tan\theta$, so that

$$\frac{d\theta}{dx} = \frac{a}{a^2 + x^2}$$

and it follows that

$$g(x) = \frac{1}{\pi} \cdot \left| \frac{a}{a^2 + x^2} \right|$$
$$= \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2} \quad \text{for} -\infty < x < \infty$$

according to Theorem 1.

EXAMPLE 8

If $F(x)$ is the value of the distribution function of the continuous random variable X at x, find the probability density of $Y = F(X)$.

Solution
As can be seen from Figure 6, the value of Y corresponding to any particular value of X is given by the area under the curve, that is, the area under the graph of the density of X to the left of x. Differentiating $y = F(x)$ with respect to x, we get

$$\frac{dy}{dx} = F'(x) = f(x)$$

and hence

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{f(x)}$$

provided $f(x) \neq 0$. It follows from Theorem 1 that

$$g(y) = f(x) \cdot \left| \frac{1}{f(x)} \right| = 1$$

for $0 < y < 1$, and we can say that y has the uniform density with $\alpha = 0$ and $\beta = 1$.

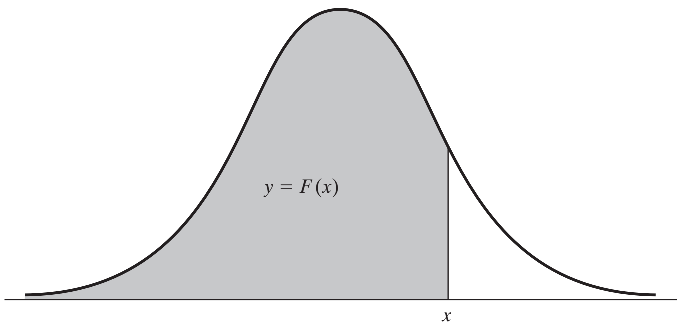

Figure 6. Diagram for Example 8.

The transformation that we performed in this example is called the **probability integral transformation**. Not only is the result of theoretical importance, but it facilitates the **simulation** of observed values of continuous random variables. A reference to how this is done, especially in connection with the normal distribution, is given in the end of the chapter.

When the conditions underlying Theorem 1 are not met, we can be in serious difficulties, and we may have to use the method of Section 2 or a generalization of Theorem 1 referred to among the references at the end of the chapter; sometimes, there is an easy way out, as in the following example.

EXAMPLE 9

If X has the standard normal distribution, find the probability density of $Z = X^2$.

Solution

Since the function given by $z = x^2$ is decreasing for negative values of x and increasing for positive values of x, the conditions of Theorem 1 are not met. However, the transformation from X to Z can be made in two steps: First, we find the probability density of $Y = |X|$, and then we find the probability density of $Z = Y^2 (= X^2)$.

As far as the first step is concerned, we already studied the transformation $Y = |X|$ in Example 2; in fact, we showed that if X has the standard normal distribution, then $Y = |X|$ has the probability density

$$g(y) = 2n(y; 0, 1) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

for $y > 0$, and $g(y) = 0$ elsewhere. For the second step, the function given by $z = y^2$ is increasing for $y > 0$, that is, for all values of Y for which $g(y) \neq 0$. Thus, we can use Theorem 1, and since

$$\frac{dy}{dz} = \frac{1}{2} z^{-\frac{1}{2}}$$

we get

$$\begin{aligned} h(z) &= \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}z} \left| \frac{1}{2} z^{-\frac{1}{2}} \right| \\ &= \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{1}{2}z} \end{aligned}$$

for $z > 0$, and $h(z) = 0$ elsewhere. Observe that since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, the distribution we have arrived at for Z is a chi-square distribution with $\nu = 1$.

---

4 Transformation Technique: Several Variables

The method of the preceding section can also be used to find the distribution of a random variable that is a function of two or more random variables. Suppose, for instance, that we are given the joint distribution of two random variables X_1 and X_2 and that we want to determine the probability distribution or the probability density of the random variable $Y = u(X_1, X_2)$. If the relationship between y and x_1 with x_2 held constant or the relationship between y and x_2 with x_1 held constant permits, we can proceed in the discrete case as in Example 4 to find the joint distribution of Y and X_2 or that of X_1 and Y and then sum on the values of the other random variable to get the marginal distribution of Y. In the continuous case, we first use Theorem 1 with the transformation formula written as

$$g(y, x_2) = f(x_1, x_2) \cdot \left| \frac{\partial x_1}{\partial y} \right|$$

or as

$$g(x_1, y) = f(x_1, x_2) \cdot \left| \frac{\partial x_2}{\partial y} \right|$$

where $f(x_1, x_2)$ and the partial derivative must be expressed in terms of y and x_2 or x_1 and y. Then we integrate out the other variable to get the marginal density of Y.

EXAMPLE 10

If X_1 and X_2 are independent random variables having Poisson distributions with the parameters λ_1 and λ_2, find the probability distribution of the random variable $Y = X_1 + X_2$.

Solution

Since X_1 and X_2 are independent, their joint distribution is given by

$$\begin{aligned} f(x_1, x_2) &= \frac{e^{-\lambda_1}(\lambda_1)^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}(\lambda_2)^{x_2}}{x_2!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1)^{x_1}(\lambda_2)^{x_2}}{x_1!x_2!} \end{aligned}$$

for $x_1 = 0, 1, 2, \ldots$ and $x_2 = 0, 1, 2, \ldots$. Since $y = x_1 + x_2$ and hence $x_1 = y - x_2$, we can substitute $y - x_2$ for x_1, getting

$$g(y, x_2) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2}}{x_2!(y-x_2)!}$$

for $y = 0, 1, 2, \ldots$ and $x_2 = 0, 1, \ldots, y$, for the joint distribution of Y and X_2. Then, summing on x_2 from 0 to y, we get

$$\begin{aligned} h(y) &= \sum_{x_2=0}^{y} \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2}}{x_2!(y-x_2)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \cdot \sum_{x_2=0}^{y} \frac{y!}{x_2!(y-x_2)!}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2} \end{aligned}$$

after factoring out $e^{-(\lambda_1+\lambda_2)}$ and multiplying and dividing by $y!$. Identifying the summation at which we arrived as the binomial expansion of $(\lambda_1+\lambda_2)^y$, we finally get

$$h(y) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^y}{y!} \quad \text{for } y = 0, 1, 2, \ldots$$

and we have thus shown that the sum of two independent random variables having Poisson distributions with the parameters λ_1 and λ_2 has a Poisson distribution with the parameter $\lambda = \lambda_1 + \lambda_2$.

EXAMPLE 11

If the joint probability density of X_1 and X_2 is given by

$$f(x_1, x_2) = \begin{cases} e^{-(x_1+x_2)} & \text{for } x_1 > 0, x_2 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability density of $Y = \dfrac{X_1}{X_1+X_2}$.

Solution

Since y decreases when x_2 increases and x_1 is held constant, we can use Theorem 1 to find the joint density of X_1 and Y. Since $y = \dfrac{x_1}{x_1+x_2}$ yields $x_2 = x_1 \cdot \dfrac{1-y}{y}$ and hence

$$\frac{\partial x_2}{\partial y} = -\frac{x_1}{y^2}$$

it follows that

$$g(x_1, y) = e^{-x_1/y}\left|-\frac{x_1}{y^2}\right| = \frac{x_1}{y^2} \cdot e^{-x_1/y}$$

for $x_1 > 0$ and $0 < y < 1$. Finally, integrating out x_1 and changing the variable of integration to $u = x_1/y$, we get

$$\begin{aligned} h(y) &= \int_0^\infty \frac{x_1}{y^2} \cdot e^{-x_1/y} dx_1 \\ &= \int_0^\infty u \cdot e^{-u} du \\ &= \Gamma(2) \\ &= 1 \end{aligned}$$

for $0 < y < 1$, and $h(y) = 0$ elsewhere. Thus, the random variable Y has the uniform density with $\alpha = 0$ and $\beta = 1$. (Note that in Exercise 7 the reader was asked to show this by the distribution function technique.)

The preceding example could also have been worked by a general method where we begin with the joint distribution of two random variables X_1 and X_2 and determine the joint distribution of two new random variables $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$. Then we can find the marginal distribution of Y_1 or Y_2 by summation or integration.

This method is used mainly in the continuous case, where we need the following theorem, which is a direct generalization of Theorem 1.

THEOREM 2. Let $f(x_1,x_2)$ be the value of the joint probability density of the continuous random variables X_1 and X_2 at (x_1,x_2). If the functions given by $y_1 = u_1(x_1,x_2)$ and $y_2 = u_2(x_1,x_2)$ are partially differentiable with respect to both x_1 and x_2 and represent a one-to-one transformation for all values within the range of X_1 and X_2 for which $f(x_1,x_2)\neq 0$, then, for these values of x_1 and x_2, the equations $y_1 = u_1(x_1,x_2)$ and $y_2 = u_2(x_1,x_2)$ can be uniquely solved for x_1 and x_2 to give $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$, and for the corresponding values of y_1 and y_2, the joint probability density of $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ is given by

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)] \cdot |J|$$

Here, J, called the **Jacobian** of the transformation, is the determinant

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Elsewhere, $g(y_1, y_2) = 0$.

We shall not prove this theorem, but information about Jacobians and their applications can be found in most textbooks on advanced calculus. There they are used mainly in connection with multiple integrals, say, when we want to change from rectangular coordinates to polar coordinates or from rectangular coordinates to spherical coordinates.

EXAMPLE 12

With reference to the random variables X_1 and X_2 of Example 11, find

(a) the joint density of $Y_1 = X_1 + X_2$ and $Y_2 = \dfrac{X_1}{X_1 + X_2}$;

(b) the marginal density of Y_2.

Solution

(a) Solving $y_1 = x_1 + x_2$ and $y_2 = \dfrac{x_1}{x_1 + x_2}$ for x_1 and x_2, we get $x_1 = y_1 y_2$ and $x_2 = y_1(1 - y_2)$, and it follows that

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1$$

Since the transformation is one-to-one, mapping the region $x_1 > 0$ and $x_2 > 0$ in the x_1x_2-plane into the region $y_1 > 0$ and $0 < y_2 < 1$ in the y_1y_2-plane, we can use Theorem 2 and it follows that

$$g(y_1, y_2) = e^{-y_1}|-y_1| = y_1e^{-y_1}$$

for $y_1 > 0$ and $0 < y_2 < 1$; elsewhere, $g(y_1, y_2) = 0$.

(b) Using the joint density obtained in part (a) and integrating out y_1, we get

$$\begin{aligned} h(y_2) &= \int_0^{\infty} g(y_1, y_2)\, dy_1 \\ &= \int_0^{\infty} y_1 e^{-y_1}\, dy_1 \\ &= \Gamma(2) \\ &= 1 \end{aligned}$$

for $0 < y_2 < 1$; elsewhere, $h(y_2) = 0$.

EXAMPLE 13

If the joint density of X_1 and X_2 is given by

$$f(x_1, x_2) = \begin{cases} 1 & \text{for } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find

(a) the joint density of $Y = X_1 + X_2$ and $Z = X_2$;

(b) the marginal density of Y.

Note that in Exercise 6 the reader was asked to work the same problem by the distribution function technique.

Solution

(a) Solving $y = x_1 + x_2$ and $z = x_2$ for x_1 and x_2, we get $x_1 = y - z$ and $x_2 = z$, so that

$$J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

Because this transformation is one-to-one, mapping the region $0 < x_1 < 1$ and $0 < x_2 < 1$ in the x_1x_2-plane into the region $z < y < z+1$ and $0 < z < 1$ in the yz-plane (see Figure 7), we can use Theorem 2 and we get

$$g(y, z) = 1 \cdot |1| = 1$$

for $z < y < z+1$ and $0 < z < 1$; elsewhere, $g(y, z) = 0$.

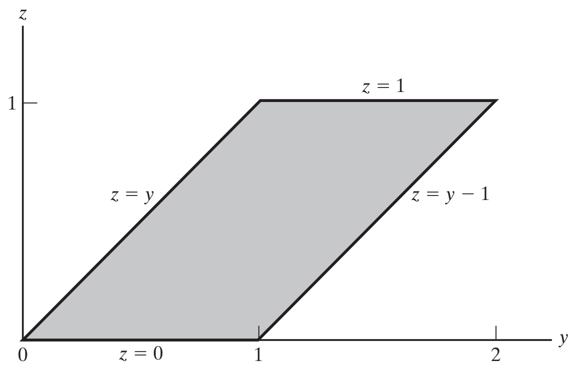

Figure 7. Transformed sample space for Example 13.

(b) Integrating out z separately for $y \leqq 0$, $0 < y < 1$, $1 < y < 2$, and $y \geqq 2$, we get

$$h(y) = \begin{cases} 0 & \text{for } y \leqq 0 \\ \displaystyle\int_0^y 1 \cdot dz = y & \text{for } 0 < y < 1 \\ \displaystyle\int_{y-1}^1 1 \cdot dz = 2 - y & \text{for } 1 < y < 2 \\ 0 & \text{for } y \geqq 2 \end{cases}$$

and to make the density function continuous, we let $h(1) = 1$. We have thus shown that the sum of the given random variables has the **triangular probability density** whose graph is shown in Figure 8.

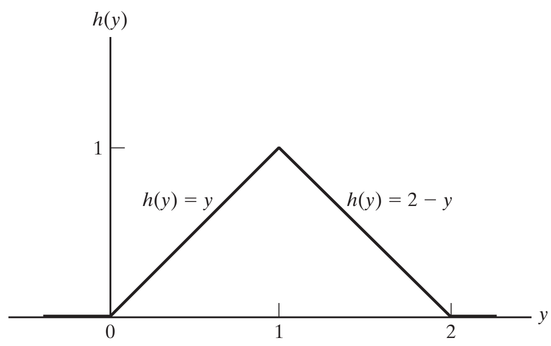

Figure 8. Triangular probability density.

So far we have considered here only functions of two random variables, but the method based on Theorem 2 can easily be generalized to functions of three or more random variables. For instance, if we are given the joint probability density of three random variables X_1, X_2, and X_3 and we want to find the joint probability density of the random variables $Y_1 = u_1(X_1, X_2, X_3)$, $Y_2 = u_2(X_1, X_2, X_3)$, and

$Y_3 = u_3(X_1, X_2, X_3)$, the general approach is the same, but the Jacobian is now the 3×3 determinant

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \dfrac{\partial x_1}{\partial y_3} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \dfrac{\partial x_2}{\partial y_3} \\ \dfrac{\partial x_3}{\partial y_1} & \dfrac{\partial x_3}{\partial y_2} & \dfrac{\partial x_3}{\partial y_3} \end{vmatrix}$$

Once we have determined the joint probability density of the three new random variables, we can find the marginal density of any two of the random variables, or any one, by integration.

EXAMPLE 14

If the joint probability density of X_1, X_2, and X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} e^{-(x_1+x_2+x_3)} & \text{for } x_1 > 0, x_2 > 0, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find

(a) the joint density of $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_2$, and $Y_3 = X_3$;

(b) the marginal density of Y_1.

Solution

(a) Solving the system of equations $y_1 = x_1 + x_2 + x_3$, $y_2 = x_2$, and $y_3 = x_3$ for x_1, x_2, and x_3, we get $x_1 = y_1 - y_2 - y_3$, $x_2 = y_2$, and $x_3 = y_3$. It follows that

$$J = \begin{vmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

and, since the transformation is one-to-one, that

$$g(y_1, y_2, y_3) = e^{-y_1} \cdot |1| \\ = e^{-y_1}$$

for $y_2 > 0$, $y_3 > 0$, and $y_1 > y_2 + y_3$; elsewhere, $g(y_1, y_2, y_3) = 0$.

(b) Integrating out y_2 and y_3, we get

$$h(y_1) = \int_0^{y_1} \int_0^{y_1 - y_3} e^{-y_1}\, dy_2\, dy_3 \\ = \frac{1}{2} y_1^2 \cdot e^{-y_1}$$

for $y_1 > 0$; $h(y_1) = 0$ elsewhere. Observe that we have shown that the sum of three independent random variables having the gamma distribution with $\alpha = 1$ and $\beta = 1$ is a random variable having the gamma distribution with $\alpha = 3$ and $\beta = 1$.

As the reader will find in Exercise 39, it would have been easier to obtain the result of part (b) of Example 14 by using the method based on Theorem 1.

Exercises

9. If X has a hypergeometric distribution with $M = 3$, $N = 6$, and $n = 2$, find the probability distribution of Y, the number of successes minus the number of failures.

10. With reference to Exercise 9, find the probability distribution of the random variable $Z = (X - 1)^2$.

11. If X has a binomial distribution with $n = 3$ and $\theta = \frac{1}{3}$, find the probability distributions of

(a) $Y = \dfrac{X}{1+X}$;

(b) $U = (X - 1)^4$.

12. If X has a geometric distribution with $\theta = \frac{1}{3}$, find the formula for the probability distribution of the random variable $Y = 4 - 5X$.

13. This question has been intentionally omitted for this edition.

14. This question has been intentionally omitted for this edition.

15. Use the transformation technique to rework Exercise 2.

16. If the probability density of X is given by

$$f(x) = \begin{cases} \dfrac{kx^3}{(1+2x)^6} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where k is an appropriate constant, find the probability density of the random variable $Y = \dfrac{2X}{1+2X}$. Identify the distribution of Y, and thus determine the value of k.

17. If the probability density of X is given by

$$f(x) = \begin{cases} \dfrac{x}{2} & \text{for } 0 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find the probability density of $Y = X^3$. Also, plot the graphs of the probability densities of X and Y and indicate the respective areas under the curves that represent $P(\frac{1}{2} < X < 1)$ and $P(\frac{1}{8} < Y < 1)$.

18. If X has a uniform density with $\alpha = 0$ and $\beta = 1$, show that the random variable $Y = -2$. In X has a gamma distribution. What are its parameters?

19. This question has been intentionally omitted for this edition.

20. Consider the random variable X with the probability density

$$f(x) = \begin{cases} \dfrac{3x^2}{2} & \text{for } -1 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Use the result of Example 2 to find the probability density of $Y = |X|$.

(b) Find the probability density of $Z = X^2 (= Y^2)$.

21. Consider the random variable X with the uniform density having $\alpha = 1$ and $\beta = 3$.

(a) Use the result of Example 2 to find the probability density of $Y = |X|$.

(b) Find the probability density of $Z = X^4 (= Y^4)$.

22. If the joint probability distribution of X_1 and X_2 is given by

$$f(x_1, x_2) = \frac{x_1 x_2}{36}$$

for $x_1 = 1, 2, 3$ and $x_2 = 1, 2, 3$, find

(a) the probability distribution of $X_1 X_2$;

(b) the probability distribution of X_1/X_2.

23. With reference to Exercise 22, find

(a) the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$;

(b) the marginal distribution of Y_1.

24. If the joint probability distribution of X and Y is given by

$$f(x, y) = \frac{(x - y)^2}{7}$$

for $x = 1, 2$ and $y = 1, 2, 3$, find

(a) the joint distribution of $U = X + Y$ and $V = X - Y$;

(b) the marginal distribution of U.

25. If X_1, X_2, and X_3 have the multinomial distribution with $n = 2$, $\theta_1 = \frac{1}{4}$, $\theta_2 = \frac{1}{3}$, and $\theta_3 = \frac{5}{12}$, find the joint probability distribution of $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$, and $Y_3 = X_3$.

26. This question has been intentionally omitted for this edition.

27. If X_1 and X_2 are independent random variables having binomial distributions with the respective parameters n_1 and θ and n_2 and θ, show that $Y = X_1 + X_2$ has the binomial distribution with the parameters $n_1 + n_2$ and θ.

28. If X_1 and X_2 are independent random variables having the geometric distribution with the parameter θ, show that $Y = X_1 + X_2$ is a random variable having the negative binomial distribution with the parameters θ and $k = 2$.

29. If X and Y are independent random variables having the standard normal distribution, show that the random variable $Z = X + Y$ is also normally distributed. (*Hint*: Complete the square in the exponent.) What are the mean and the variance of this normal distribution?

30. Consider two random variables X and Y with the joint probability density

$$f(x,y) = \begin{cases} 12xy(1-y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability density of $Z = XY^2$ by using Theorem 1 to determine the joint probability density of Y and Z and then integrating out y.

31. Rework Exercise 30 by using Theorem 2 to determine the joint probability density of $Z = XY^2$ and $U = Y$ and then finding the marginal density of Z.

32. Consider two independent random variables X_1 and X_2 having the same Cauchy distribution

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{for } -\infty < x < \infty$$

Find the probability density of $Y_1 = X_1 + X_2$ by using Theorem 1 to determine the joint probability density of X_1 and Y_1 and then integrating out x_1. Also, identify the distribution of Y_1.

33. Rework Exercise 32 by using Theorem 2 to determine the joint probability density of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ and then finding the marginal density of Y_1.

34. Consider two random variables X and Y whose joint probability density is given by

$$f(x,y) = \begin{cases} \frac{1}{2} & \text{for } x > 0, y > 0, x + y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability density of $U = Y - X$ by using Theorem 1.

35. Rework Exercise 34 by using Theorem 2 to determine the joint probability density of $U = Y - X$ and $V = X$ and then finding the marginal density of U.

36. Let X_1 and X_2 be two continuous random variables having the joint probability density

$$f(x_1,x_2) = \begin{cases} 4x_1x_2 & \text{for } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the joint probability density of $Y_1 = X_1^2$ and $Y_2 = X_1X_2$.

37. Let X and Y be two continuous random variables having the joint probability density

$$f(x,y) = \begin{cases} 24xy & \text{for } 0 < x < 1, 0 < y < 1, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the joint probability density of $Z = X + Y$ and $W = X$.

38. Let X and Y be two independent random variables having identical gamma distributions.
(a) Find the joint probability density of the random variables $U = \dfrac{X}{X+Y}$ and $V = X + Y$.
(b) Find and identify the marginal density of U.

39. The method of transformation based on Theorem 1 can be generalized so that it applies also to random variables that are functions of two or more random variables. So far we have used this method only for functions of two random variables, but when there are three, for example, we introduce the new random variable in place of one of the original random variables, and then we eliminate (by summation or integration) the other two random variables with which we began. Use this method to rework Example 14.

40. In Example 13 we found the probability density of the sum of two independent random variables having the uniform density with $\alpha = 0$ and $\beta = 1$. Given a third random variable X_3, which has the same uniform density and is independent of both X_1 and X_2, show that if $U = Y + X_3 = X_1 + X_2 + X_3$, then
(a) the joint probability density of U and Y is given by

$$g(u,y) = \begin{cases} y & \text{for Regions I and II of Figure 9} \\ 2-y & \text{for Regions III and IV of Figure 9} \\ 0 & \text{elsewhere} \end{cases}$$

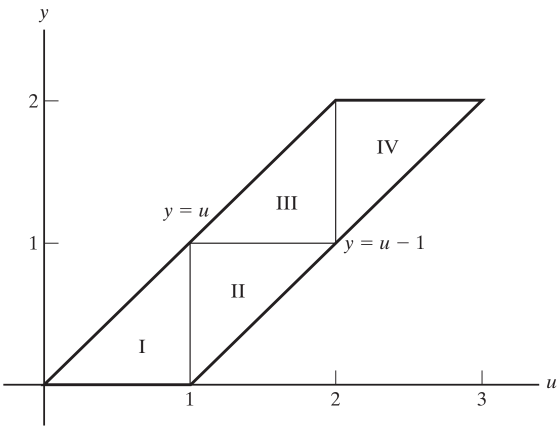

Figure 9. Diagram for Exercise 40.

(b) the probability density of U is given by

$$h(u) = \begin{cases} 0 & \text{for } u \leqq 0 \\ \frac{1}{2}u^2 & \text{for } 0 < u < 1 \\ \frac{1}{2}u^2 - \frac{3}{2}(u-1)^2 & \text{for } 1 < u < 2 \\ \frac{1}{2}u^2 - \frac{3}{2}(u-1)^2 + \frac{3}{2}(u-2)^2 & \text{for } 2 < u < 3 \\ 0 & \text{for } u \geqq 3 \end{cases}$$

Note that if we let $h(1) = h(2) = \frac{1}{2}$, this will make the probability density of U continuous.

5 Moment-Generating Function Technique

Moment-generating functions can play an important role in determining the probability distribution or density of a function of random variables when the function is a linear combination of n *independent* random variables. We shall illustrate this technique here when such a linear combination is, in fact, the sum of n independent random variables, leaving it to the reader to generalize it in Exercises 45 and 46.

The method is based on the following theorem that the moment-generating function of the sum of n independent random variables equals the product of their moment-generating functions.

THEOREM 3. If $X_1, X_2, \ldots$, and X_n are independent random variables and $Y = X_1 + X_2 + \cdots + X_n$, then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

where $M_{X_i}(t)$ is the value of the moment-generating function of X_i at t.

Proof Making use of the fact that the random variables are independent and hence

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n)$$

according to the following definition "**INDEPENDENCE OF DISCRETE RANDOM VARIABLES**. *If* $\mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$ *is the value of the joint probability distribution of the discrete random variables* $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{n}$ *at* $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$ *and* $\mathrm{f}_\mathrm{i}(\mathrm{x}_\mathrm{i})$ *is the value of the marginal distribution of* X_i *at* x_i *for* $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$*, then the* n *random variables are* ***independent*** *if and only if* $f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n)$ *for all* $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$ *within their range*", we can write

$$
\begin{aligned}
M_Y(t) &= E(e^{Yt}) \\
&= E\left[e^{(X_1+X_2+\cdots+X_n)t}\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{(x_1+x_2+\cdots+x_n)t} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \\
&= \int_{-\infty}^{\infty} e^{x_1 t} f_1(x_1)\, dx_1 \cdot \int_{-\infty}^{\infty} e^{x_2 t} f_2(x_2)\, dx_2 \cdots \int_{-\infty}^{\infty} e^{x_n t} f_n(x_n)\, dx_n \\
&= \prod_{i=1}^{n} M_{X_i}(t)
\end{aligned}
$$

which proves the theorem for the continuous case. To prove it for the discrete case, we have only to replace all the integrals by sums.

Note that if we want to use Theorem 3 to find the probability distribution or the probability density of the random variable $Y = X_1 + X_2 + \cdots + X_n$, we must be able to identify whatever probability distribution or density corresponds to $M_Y(t)$.

EXAMPLE 15

Find the probability distribution of the sum of n independent random variables X_1, $X_2, \ldots, X_n$ having Poisson distributions with the respective parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Solution

By the theorem "The moment-generating function of the Poisson distribution is given by $M_X(t) = e^{\lambda(e^t-1)}$" we have

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)}$$

hence, for $Y = X_1 + X_2 + \cdots + X_n$, we obtain

$$M_Y(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{(\lambda_1+\lambda_2+\cdots+\lambda_n)(e^t-1)}$$

which can readily be identified as the moment-generating function of the Poisson distribution with the parameter $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Thus, the distribution of the sum of n independent random variables having Poisson distributions with the parameters λ_i is a Poisson distribution with the parameter $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Note that in Example 10 we proved this for $n = 2$.

EXAMPLE 16

If $X_1, X_2, \ldots, X_n$ are independent random variables having exponential distributions with the same parameter θ, find the probability density of the random variable $Y = X_1 + X_2 + \cdots + X_n$.

Solution

Since the exponential distribution is a gamma distribution with $\alpha = 1$ and $\beta = \theta$, we have

$$M_{X_i}(t) = (1 - \theta t)^{-1}$$

and hence

$$M_Y(t) = \prod_{i=1}^{n}(1-\theta t)^{-1} = (1-\theta t)^{-n}$$

Identifying the moment-generating function of Y as that of a gamma distribution with $\alpha = n$ and $\beta = \theta$, we conclude that the distribution of the sum of n independent random variables having exponential distributions with the same parameter θ is a gamma distribution with the parameters $\alpha = n$ and $\beta = \theta$. Note that this agrees with the result of Example 14, where we showed that the sum of three independent random variables having exponential distributions with the parameter $\theta = 1$ has a gamma distribution with $\alpha = 3$ and $\beta = 1$.

Theorem 3 also provides an easy and elegant way of deriving the moment-generating function of the binomial distribution. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables having the same Bernoulli distribution $f(x;\theta) = \theta^x(1-\theta)^{1-x}$ for $x = 0, 1$. We have

$$M_{X_i}(t) = e^{0\cdot t}(1-\theta) + e^{1\cdot t}\theta = 1+\theta(e^t-1)$$

so that Theorem 3 yields

$$M_Y(t) = \prod_{i=1}^{n}[1+\theta(e^t-1)] = [1+\theta(e^t-1)]^n$$

This moment-generating function is readily identified as that of the binomial distribution with the parameters n and θ. Of course, $Y = X_1 + X_2 + \cdots + X_n$ is the total number of successes in n trials, since X_1 is the number of successes on the first trial, X_2 is the number of successes on the second trial, ..., and X_n is the number of successes on the nth trial. This is a fruitful way of looking at the binomial distribution.

Exercises

41. Use the moment-generating function technique to rework Exercise 27.

42. Find the moment-generating function of the negative binomial distribution by making use of the fact that if k independent random variables have geometric distributions with the same parameter θ, their sum is a random variable having the negative binomial distribution with the parameters θ and k.

43. If n independent random variables have the same gamma distribution with the parameters α and β, find the moment-generating function of their sum and, if possible, identify its distribution.

44. If n independent random variables X_i have normal distributions with the means μ_i and the standard deviations σ_i, find the moment-generating function of their sum and identify the corresponding distribution, its mean, and its variance.

45. Prove the following generalization of Theorem 3: If $X_1, X_2, \ldots$, and X_n are independent random variables and $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

where $M_{X_i}(t)$ is the value of the moment-generating function of X_i at t.

46. Use the result of Exercise 45 to show that, if n independent random variables X_i have normal distributions with the means μ_i and the standard deviations σ_i, then $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ has a normal distribution. What are the mean and the variance of this distribution?

6 The Theory in Application

Examples of the need for transformations in solving practical problems abound. To illustrate these applications, we give three examples. The first example illustrates an application of the transformation technique to a simple problem in electrical engineering.

EXAMPLE 17

Suppose the resistance in a simple circuit varies randomly in response to environmental conditions. To determine the effect of this variation on the current flowing through the circuit, an experiment was performed in which the resistance (R) was varied with equal probabilities on the interval $0 < R \leq A$ and the ensuing voltage (E) was measured. Find the distribution of the random variable I, the current flowing through the circuit.

Solution

Using the well-known relation $E = IR$, we have $I = u(R) = \dfrac{E}{R}$. The probability distribution of R is given by $f(R) = \dfrac{1}{A}$ for $0 < R \leq A$. Thus, $w(I) = \dfrac{E}{I}$, and the probability density of I is given by

$$g(I) = f(R) \cdot |w'(I)| = \frac{1}{A}\left|-\frac{E}{R^2}\right| = \frac{E}{AR^2} \qquad R > 0$$

It should be noted, with respect to this example, that this is a designed experiment in as much as the distribution of R was preselected as a uniform distribution. If the nominal value of R is to be the mean of this distribution, some other distribution might have been selected to impart better properties to this estimate.

The next example illustrates transformations of data to normality.

EXAMPLE 18

What underlying distribution of the data is assumed when the square-root transformation is used to obtain approximately normally distributed data? (Assume the data are nonnegative, that is, the probability of a negative observation is zero.)

Solution

A simple alternate way to use the distribution-function technique is to write down the differential element of the density function, $f(x)\,dx$, of the transformed observations, y, and to substitute x^2 for y. (When we do this, we must remember that the differential element, dy, must be changed to $dx = 2x\,dx$.) We obtain

$$f(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma} \cdot 2x \cdot e^{-\frac{1}{2}(x^2-\mu)^2/\sigma^2}\,dx$$

The required density function is given by

$$f(x) = \sqrt{\frac{2}{\pi\sigma^2}}\,xe^{-\frac{1}{2}(x^2-\mu)^2/\sigma^2}$$

This distribution is not immediately recognizable, but it can be graphed quickly using appropriate computer software.

The final example illustrates an application to waiting-time problems.

EXAMPLE 19

Let us assume that the decay of a radioactive element is exponentially distributed, so that $f(x) = \lambda e^{-\lambda x}$ for $\lambda > 0$ and $x > 0$; that is, the time for the nucleus to emit the first α particle is x (in seconds). It can be shown that such a process has no memory; that is, the time *between successive emissions* also can be described by this distribution. Thus, it follows that successive emissions of α particles are independent. If the parameter λ equals 5, find the probability that a given substance will emit 2 particles in less than or equal to 3 seconds.

Solution

Let x_i be the waiting time between emissions i and $i+1$, for $i = 0, 1, 2, \ldots, n-1$. Then the total time for n emissions to take place is the sum $T = x_0 + x_1 + \cdots + x_{n-1}$. The moment-generating function of this sum is given in Example 16 to be

$$M_T(t) = (1 - t/\lambda)^{-n}$$

This can be recognized as the moment-generating function of the gamma distribution with parameters $\alpha = n = 2$ and $\beta = 1/\lambda = 1/5$. The required probability is given by

$$P\left(T \leq 3; \alpha = 10, \beta = \frac{1}{5}\right) = \frac{1}{\frac{1}{5}\Gamma(2)} \int_0^3 x\, e^{-5x} dx$$

Integrating by parts, the integral becomes

$$P(T \leq 3) = -\frac{1}{5} x e^{-5x}\Big|_0^3 - \int_0^3 -\frac{1}{5} e^{-5x}\, dx = 1 - 1.6e^{-15}$$

Without further evaluation, it is clear that this event is virtually certain to occur.

Applied Exercises — SECS. 1–2

47. This question has been intentionally omitted for this edition.

48. This question has been intentionally omitted for this edition.

49. This question has been intentionally omitted for this edition.

50. Let X be the amount of premium gasoline (in 1,000 gallons) that a service station has in its tanks at the beginning of a day, and Y the amount that the service station sells during that day. If the joint density of X and Y is given by

$$f(x, y) = \begin{cases} \dfrac{1}{200} & \text{for } 0 < y < x < 20 \\ 0 & \text{elsewhere} \end{cases}$$

use the distribution function technique to find the probability density of the amount that the service station has left in its tanks at the end of the day.

51. The percentages of copper and iron in a certain kind of ore are, respectively, X_1 and X_2. If the joint density of these two random variables is given by

$$f(x_1, x_2) = \begin{cases} \dfrac{3}{11}(5x_1 + x_2) & \text{for } x_1 > 0, x_2 > 0, \\ & \text{and } x_1 + 2x_2 < 2 \\ 0 & \text{elsewhere} \end{cases}$$

use the distribution function technique to find the probability density of $Y = X_1 + X_2$. Also find $E(Y)$, the expected total percentage of copper and iron in the ore.

SECS. 3–4

52. According to the Maxwell–Boltzmann law of theoretical physics, the probability density of V, the velocity of a gas molecule, is

$$f(v) = \begin{cases} kv^2 e^{-\beta v^2} & \text{for } v > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where β depends on its mass and the absolute temperature and k is an appropriate constant. Show that the kinetic energy $E = \frac{1}{2}mV^2$, where m the mass of the molecule is a random variable having a gamma distribution.

53. This question has been intentionally omitted for this edition.

54. This question has been intentionally omitted for this edition.

55. This question has been intentionally omitted for this edition.

56. Use a computer program to generate 10 "pseudorandom" numbers having the standard normal distribution.

57. Describe how the probability integral transformation might have been used by the writers of the software that you used to produce the result of Exercise 56.

SEC. 5

58. A lawyer has an unlisted number on which she receives on the average 2.1 calls every half-hour and a listed number on which she receives on the average 10.9 calls every half-hour. If it can be assumed that the numbers of calls that she receives on these phones are independent random variables having Poisson distributions, what are the probabilities that in half an hour she will receive altogether

(a) 14 calls;

(b) at most 6 calls?

59. In a newspaper ad, a car dealer lists a 2001 Chrysler, a 2010 Ford, and a 2008 Buick. If the numbers of inquiries he will get about these cars may be regarded as independent random variables having Poisson distributions with the parameters $\lambda_1 = 3.6$, $\lambda_2 = 5.8$, and $\lambda_3 = 4.6$, what are the probabilities that altogether he will receive

(a) fewer than 10 inquiries about these cars;

(b) anywhere from 15 to 20 inquiries about these cars;

(c) at least 18 inquiries about these cars?

60. With reference to Exercise 59, what is the probability that the car dealer will receive six inquiries about the 2010 Ford and eight inquiries about the other two cars?

61. If the number of complaints a dry-cleaning establishment receives per day is a random variable having the Poisson distribution with $\lambda = 3.3$, what are the probabilities that it will receive

(a) 2 complaints on any given day;

(b) 5 complaints altogether on any two given days;

(c) at least 12 complaints altogether on any three given days?

62. The number of fish that a person catches per hour at Woods Canyon Lake is a random variable having the Poisson distribution with $\lambda = 1.6$. What are the probabilities that a person fishing there will catch

(a) four fish in 2 hours;

(b) at least two fish in 3 hours;

(c) at most three fish in 4 hours?

63. If the number of minutes it takes a service station attendant to balance a tire is a random variable having an exponential distribution with the parameter $\theta = 5$, what are the probabilities that the attendant will take

(a) less than 8 minutes to balance two tires;

(b) at least 12 minutes to balance three tires?

64. If the number of minutes that a doctor spends with a patient is a random variable having an exponential distribution with the parameter $\theta = 9$, what are the probabilities that it will take the doctor at least 20 minutes to treat

(a) one patient; **(b)** two patients; **(c)** three patients?

65. If X is the number of 7's obtained when rolling a pair of dice three times, find the probability that $Y = X^2$ will exceed 2.

66. If X has the exponential distribution given by $f(x) = 0.5\,e^{-0.5x}$, $x > 0$, find the probability that $x > 1$.

SEC. 6

67. If, d, the diameter of a circle is selected at random from the density function

$$f(d) = k\left(1 - \frac{d}{5}\right), 0 < d < 5,$$

(a) find the value of k so that *f(d)* is a probability density;

(b) find the density function of the areas of the circles so selected.

68. Show that the underlying distribution function of Example 18 is, indeed, a probability distribution, and use a computer program to graph the density function.

69. If $X = \ln Y$ has a normal distribution with the mean μ and the standard deviation σ, find the probability density of Y which is said to have the **log-normal** distribution.

70. The logarithm of the ratio of the output to the input current of a transistor is called its current gain. If current gain measurements made on a certain transistor are normally distributed with $\mu = 1.8$ and $\sigma = 0.05$, find the probability that the current gain will exceed the required minimum value of 6.0.

References

The use of the probability integral transformation in problems of simulation is discussed in

Johnson, R. A., *Miller and Freund's Probability and Statistics for Engineers*, 6th ed. Upper Saddle River, N.J.: Prentice Hall, 2000.

A generalization of Theorem 1, which applies when the interval within the range of X for which $f(x) \neq 0$ can be partitioned into k subintervals so that the conditions of Theorem 1 apply separately for each of the subintervals, may be found in

Walpole, R. E., and Myers, R. H., *Probability and Statistics for Engineers and Scientists*, 4th ed. New York: Macmillan Publishing Company, Inc., 1989.

More detailed and more advanced treatments of the material in this chapter are given in many advanced texts on mathematical statistics; for instance, in

Hogg, R. V., and Craig, A. T., *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan Publishing Company, Inc., 1978,

Roussas, G. G., *A First Course in Mathematical Statistics*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1973,

Wilks, S. S., *Mathematical Statistics*. New York: John Wiley & Sons, Inc., 1962.

Answers to Odd-Numbered Exercises

1 $g(y) = \frac{1}{\theta} e^y e^{-(1/\theta)} e^y$ for $-\infty < y < \infty$.

3 $g(y) = 2y$ for $0 < y < 1$ and $g(y) = 0$ elsewhere.

5 **(a)** $f(y) = \dfrac{1}{\theta_1 - \theta_2} \cdot (e^{-y/\theta_1} - e^{-y/\theta_2})$ for $y > 0$ and $f(y) = 0$ elsewhere; **(b)** $f(y) = \dfrac{1}{\theta^2} \cdot y e^{-y/\theta}$ for $y > 0$ and $f(y) = 0$ elsewhere.

9 $h(-2) = \frac{1}{5}, h(0) = \frac{3}{5}$, and $h(2) = \frac{1}{5}$.

11 **(a)** $g(0) = \frac{8}{27}$, $g(\frac{1}{2}) = \frac{12}{27}$, $g(\frac{2}{3}) = \frac{6}{27}$, $g(\frac{3}{4}) = \frac{1}{27}$;
(b) $g(0) = \frac{12}{27}$, $g(1) = \frac{14}{27}$, $g(16) = \frac{1}{27}$.

13 $g(0) = \frac{1}{3}$, $g(1) = \frac{1}{3}$, $g(2) = \frac{1}{3}$.

17 $g(y) = \dfrac{1}{6} y^{\frac{-1}{3}}$.

21 **(a)** $g(y) = \frac{1}{8} y^{-3/4}$ for $0 < y < 1$ and $g(y) = \frac{1}{4}$ for $1 < y < 3$;
(b) $h(z) = \frac{1}{16} \cdot z^{-3/4}$ for $1 < z < 81$ and $h(z) = 0$ elsewhere.

23 **(a)** $f(2, 0) = \frac{1}{36}$, $f(3, -1) = \frac{2}{36}$, $f(3, 1) = \frac{2}{36}$, $f(4, -2) = \frac{3}{36}$, $f(4, 0) = \frac{4}{36}$, $f(4, 2) = \frac{3}{36}$, $f(5, -1) = \frac{6}{36}$, $f(5, 1) = \frac{6}{36}$, and $f(6, 0) = \frac{9}{36}$;
(b) $g(2) = \frac{1}{36}$, $g(3) = \frac{4}{36}$, $g(4) = \frac{10}{36}$, $g(5) = \frac{12}{36}$, and $g(6) = \frac{9}{36}$.

25 **(b)** $g(0, 0, 2) = \frac{25}{144}$, $g(1, -1, 1) = \frac{5}{18}$, $g(1, 1, 1) = \frac{5}{24}$, $g(2, -2, 0) = \frac{1}{9}$, $g(2, 0, 0) = \frac{1}{6}$, and $g(2, 2, 0) = \frac{1}{16}$.

29 $\mu = 0$ and $\sigma^2 = 2$.

31 $g(z,u) = 12z(u^{-3} - u^{-2})$ over the region bounded by $z = 0, u = 1$, and $z = u^2$, and $g(z,u) = 0$ elsewhere; $h(z) = 6z + 6 - 12\sqrt{z}$ for $0 < z < 1$ and $h(z) = 0$ elsewhere.

33 The marginal distribution is the Cauchy distribution $g(y) = \dfrac{1}{\pi} \cdot \dfrac{2}{4 + y^2}$ for $-\infty < y < \infty$.

35 $f(u,v) = \frac{1}{2}$ over the region bounded by $v = 0, u = -v$, and $2v + u = 2$, and $f(u,v) = 0$ elsewhere; $g(u) = \frac{1}{4}(2+u)$ for $-2 < u \leq 0, g(u) = \frac{1}{4}(2-u)$ for $0 < u < 2$ and $g(u) = 0$ elsewhere.

37 $g(w, z) = 24w(z - w)$ over the region bounded by $w = 0, z = 1$, and $z = w$; $g(w,z) = 0$ elsewhere.

43 It is a gamma distribution with the parameters αn and β.

51 $g(y) = \frac{9}{11} \cdot y^2$ for $0 < y \leq 1, g(y) = \dfrac{3(2-y)(7y-4)}{11}$ for $1 < y < 2$, and $g(y) = 0$ elsewhere.

53 $h(r) = 2r$ for $0 < r < 1$ and $h(r) = 0$ elsewhere.

55 $g(v, w) = 5e^{-v}$ for $0.2 < w < 0.4$ and $v > 0$; $h(v) = e^{-v}$ for $v > 0$ and $h(v) = 0$ elsewhere.

59 **(a)** 0.1093; **(b)** 0.3817; **(c)** 0.1728.

61 **(a)** 0.2008; **(b)** 0.1420; **(c)** 0.2919.

63 **(a)** 0.475; **(b)** 0.570.

65 $\frac{2}{27}$.

67 **(a)** $\frac{2}{5}$; **(b)** $g(A) = \frac{2}{5}\left(\dfrac{1}{\sqrt{\pi}} A^{-1/2} - 1\right)$ for $0 < A < \frac{25}{4}\pi$ and $g(A) = 0$ elsewhere.

69 $g(y) = \dfrac{1}{\sqrt{2\pi}\sigma} \cdot \dfrac{1}{y} \cdot e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}$ for $y > 0$ and $g(y) = 0$ elsewhere.

Sampling Distributions

1 Introduction

Statistics concerns itself mainly with conclusions and predictions resulting from chance outcomes that occur in carefully planned experiments or investigations. Drawing such conclusions usually involves taking sample observations from a given population and using the results of the sample to make inferences about the population itself, its mean, its variance, and so forth. To do this requires that we first find the distributions of certain functions of the random variables whose values make up the sample, called **statistics**. (An example of such a statistic is the sample mean.) The properties of these distributions then allow us to make probability statements about the resulting inferences drawn from the sample about the population. The theory to be given in this chapter forms an important foundation for the theory of statistical inference.

Inasmuch as statistical inference can be loosely defined as a process of drawing conclusions from a sample about the population from which it is drawn, it is useful to have the following definition.

DEFINITION 1. POPULATION. *A set of numbers from which a sample is drawn is referred to as a **population**. The distribution of the numbers constituting a population is called the **population distribution**.*

To illustrate, suppose a scientist must choose and then weigh 5 of 40 guinea pigs as part of an experiment, a layman might say that the ones she selects constitute the sample. This is how the term "sample" is used in everyday language. In statistics, it is preferable to look upon the weights of the 5 guinea pigs as a sample from the population, which consists of the weights of all 40 guinea pigs. In this way, the population as well as the sample consists of numbers. Also, suppose that, to estimate the average useful life of a certain kind of transistor, an engineer selects 10 of these transistors, tests them over a period of time, and records for each one the time to failure. If these times to failure are values of independent random variables having an exponential distribution with the parameter θ, we say that they constitute a sample from this exponential population.

As can well be imagined, not all samples lend themselves to valid generalizations about the populations from which they came. In fact, most of the methods of inference discussed in this chapter are based on the assumption that we are dealing with

From Chapter 8 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

random samples. In practice, we often deal with random samples from populations that are finite, but large enough to be treated as if they were infinite. Thus, most statistical theory and most of the methods we shall discuss apply to samples from infinite populations, and we shall begin here with a definition of random samples from infinite populations. Random samples from finite populations will be treated later in Section 3.

DEFINITION 2. RANDOM SAMPLE. *If* $X_1, X_2, \ldots, X_n$ *are independent and identically distributed random variables, we say that they constitute a* ***random sample*** *from the infinite population given by their common distribution.*

If $f(x_1, x_2, \ldots, x_n)$ is the value of the joint distribution of such a set of random variables at $(x_1, x_2, \ldots, x_n)$, by virtue of independence we can write

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

where $f(x_i)$ is the value of the population distribution at x_i. Observe that Definition 2 and the subsequent discussion apply also to sampling with replacement from finite populations; sampling without replacement from finite populations is discussed in section 3.

Statistical inferences are usually based on **statistics**, that is, on random variables that are functions of a set of random variables $X_1, X_2, \ldots, X_n$, constituting a random sample. Typical of what we mean by "statistic" are the **sample mean** and the **sample variance**.

DEFINITION 3. SAMPLE MEAN AND SAMPLE VARIANCE. *If* $X_1, X_2, \ldots, X_n$ *constitute a random sample, then the* ***sample mean*** *is given by*

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

and the ***sample variance*** *is given by*

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

As they are given here, these definitions apply only to random samples, but the sample mean and the sample variance can, similarly, be defined for any set of random variables $X_1, X_2, \ldots, X_n$.

It is common practice also to apply the terms "random sample," "statistic," "sample mean," and "sample variance" to the values of the random variables instead of the random variables themselves. Intuitively, this makes more sense and it conforms with colloquial usage. Thus, we might calculate

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

for observed sample data and refer to these statistics as the sample mean and the sample variance. Here, the x_i, $\overline{x}$, and s^2 are values of the corresponding random

† The note has been intentionally omitted for this edition.

variables X_i, $\overline{X}$, and S^2. Indeed, the formulas for $\bar{x}$ and s^2 are used even when we deal with any kind of data, not necessarily sample data, in which case we refer to $\bar{x}$ and s^2 simply as the mean and the variance.

These, and other statistics that will be introduced in this chapter, are those mainly used in statistical inference. Sample statistics such as the sample mean and sample variance play an important role in estimating the parameters of the population from which the corresponding random samples were drawn.

2 The Sampling Distribution of the Mean

Inasmuch as the values of sampling statistics can be expected to vary from sample to sample, it is necessary that we find the distribution of such statistics. We call these distributions **sampling distributions**, and we make important use of them in determining the properties of the inferences we draw from the sample about the parameters of the population from which it is drawn.

First let us study some theory about the **sampling distribution of the mean**, making only some very general assumptions about the nature of the populations sampled.

Theorem 1. If $X_1, X_2, \ldots, X_n$ constitute a random sample from an infinite population with the mean μ and the variance σ^2, then

$$E(\overline{X}) = \mu \quad \text{and} \quad \text{var}(\overline{X}) = \frac{\sigma^2}{n}$$

Proof Letting $Y = \overline{X}$ and hence setting $a_i = \frac{1}{n}$, we get

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n} \cdot \mu = n\left(\frac{1}{n} \cdot \mu\right) = \mu$$

since $E(X_i) = \mu$. Then, by the corollary of a theorem "If the random variables $X_1, X_2, \ldots, X_n$ are independent and $Y = \sum_{i=1}^{n} a_i X_i$, then $\text{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i)$", we conclude that

$$\text{var}(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n^2} \cdot \sigma^2 = n\left(\frac{1}{n^2} \cdot \sigma^2\right) = \frac{\sigma^2}{n}$$

It is customary to write $E(\overline{X})$ as $\mu_{\overline{X}}$ and $\text{var}(\overline{X})$ as $\sigma^2_{\overline{X}}$ and refer to $\sigma_{\overline{X}}$ as the **standard error of the mean**. The formula for the standard error of the mean, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$, shows that the standard deviation of the distribution of $\overline{X}$ decreases when n, the **sample size**, is increased. This means that when n becomes larger and we actually have more information (the values of more random variables), we can expect values of $\overline{X}$ to be closer to μ, the quantity that they are intended to estimate.

Theorem 2. For any positive constant c, the probability that $\overline{X}$ will take on a value between $\mu - c$ and $\mu + c$ is at least

$$1 - \frac{\sigma^2}{nc^2}$$

When $n \to \infty$, this probability approaches 1.

This result, called a **law of large numbers**, is primarily of theoretical interest. Of much more practical value is the **central limit theorem**, one of the most important theorems of statistics, which concerns the limiting distribution of the **standardized mean** of n random variables when $n\to\infty$. We shall prove this theorem here only for the case where the n random variables are a random sample from a population whose moment-generating function exists. More general conditions under which the theorem holds are given in Exercises 7 and 9, and the most general conditions under which it holds are referred to at the end of this chapter.

THEOREM 3. CENTRAL LIMIT THEOREM. If $X_1, X_2, \ldots, X_n$ constitute a random sample from an infinite population with the mean μ, the variance σ^2, and the moment-generating function $M_X(t)$, then the limiting distribution of

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

as $n\to\infty$ is the standard normal distribution.

Proof First using the third part and then the second of the given theorem "If a and b are constants, then **1.** $M_{X+a}(t) = E[e^{(X+a)t}] = e^{at}\cdot M_X(t)$; **2.** $M_{bX}(t) = E(e^{bXt}) = M_X(bt)$; **3.** $M_{\frac{X+a}{b}}(t) = E[e^{\left(\frac{X+a}{b}\right)t}] = e^{\frac{a}{b}t}\cdot M_X\left(\frac{t}{b}\right)$", we get

$$M_Z(t) = M_{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}(t) = e^{-\sqrt{n}\,\mu t/\sigma}\cdot M_{\overline{X}}\left(\frac{\sqrt{n}t}{\sigma}\right)$$
$$= e^{-\sqrt{n}\,\mu t/\sigma}\cdot M_{n\overline{X}}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Since $n\overline{X} = X_1 + X_2 + \cdots + X_n$,

$$M_Z(t) = e^{-\sqrt{n}\,\mu t/\sigma}\cdot\left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

and hence that

$$\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n\cdot\ln M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Expanding $M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$ as a power series in t, we obtain

$$\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n\cdot\ln\left[1+\mu_1'\frac{t}{\sigma\sqrt{n}}+\mu_2'\frac{t^2}{2\sigma^2 n}+\mu_3'\frac{t^3}{6\sigma^3 n\sqrt{n}}+\cdots\right]$$

where μ_1', μ_2', and μ_3' are the moments about the origin of the population distribution, that is, those of the original random variables X_i.

If n is sufficiently large, we can use the expansion of $\ln(1+x)$ as a power series in x, getting

$$\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n\left\{\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]\right.$$

$$-\frac{1}{2}\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]^2$$

$$\left.+\frac{1}{3}\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]^3 - \cdots\right\}$$

Then, collecting powers of t, we obtain

$$\ln M_Z(t) = \left(-\frac{\sqrt{n}\,\mu}{\sigma} + \frac{\sqrt{n}\,\mu_1'}{\sigma}\right)t + \left(\frac{\mu_2'}{2\sigma^2} - \frac{\mu_1'^2}{2\sigma^2}\right)t^2$$

$$+\left(\frac{\mu_3'}{6\sigma^3\sqrt{n}} - \frac{\mu_1'\cdot\mu_2'}{2\sigma^3\sqrt{n}} + \frac{\mu_1'^3}{3\sigma^3\sqrt{n}}\right)t^3 + \cdots$$

and since $\mu_1' = \mu$ and $\mu_2' - (\mu_1')^2 = \sigma^2$, this reduces to

$$\ln M_Z(t) = \frac{1}{2}t^2 + \left(\frac{\mu_3'}{6} - \frac{\mu_1'\mu_2'}{2} + \frac{\mu_1'^3}{6}\right)\frac{t^3}{\sigma^3\sqrt{n}} + \cdots$$

Finally, observing that the coefficient of t^3 is a constant times $\dfrac{1}{\sqrt{n}}$ and in general, for $r \geqq 2$, the coefficient of t^r is a constant times $\dfrac{1}{\sqrt{n^{r-2}}}$, we get

$$\lim_{n\to\infty} \ln M_Z(t) = \frac{1}{2}t^2$$

and hence

$$\lim_{n\to\infty} M_Z(t) = e^{\frac{1}{2}t^2}$$

since the limit of a logarithm equals the logarithm of the limit (provided these limits exist). An illustration of this theorem is given in Exercise 13 and 14.

Sometimes, the central limit theorem is interpreted incorrectly as implying that the distribution of $\overline{X}$ approaches a normal distribution when $n \to \infty$. This is incorrect because $\mathrm{var}(\overline{X}) \to 0$ when $n \to \infty$; on the other hand, the central limit theorem does justify approximating the distribution of $\overline{X}$ with a normal distribution having the mean μ and the variance $\dfrac{\sigma^2}{n}$ when n is large. In practice, this approximation is used when $n \geqq 30$ regardless of the actual shape of the population sampled. For smaller values of n the approximation is questionable, but see Theorem 4.

EXAMPLE 1

A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 milliliters and a standard deviation of 15 milliliters. What is the probability that the average (mean) amount dispensed in a random sample of size 36 is at least 204 milliliters?

Solution

According to Theorem 1, the distribution of $\overline{X}$ has the mean $\mu_{\overline{X}} = 200$ and the standard deviation $\sigma_{\overline{X}} = \dfrac{15}{\sqrt{36}} = 2.5$, and according to the central limit theorem, this distribution is approximately normal. Since $z = \dfrac{204 - 200}{2.5} = 1.6$, it follows from Table III of "Statistical Tables" that $P(\overline{X} \geqq 204) = P(Z \geqq 1.6) = 0.5000 - 0.4452 = 0.0548$.

It is of interest to note that when the population we are sampling is normal, the distribution of $\overline{X}$ is a normal distribution regardless of the size of n.

THEOREM 4. If $\overline{X}$ is the mean of a random sample of size n from a normal population with the mean μ and the variance σ^2, its sampling distribution is a normal distribution with the mean μ and the variance σ^2/n.

Proof According to Theorems "If a and b are constants, then **1.** $M_{X+a}(t) = E[e^{(X+a)t}] = e^{at} \cdot M_X(t)$; **2.** $M_{bX}(t) = E(e^{bXt}) = M_X(bt)$; **3.** $M_{\frac{X+a}{b}}(t) = E[e^{\left(\frac{X+a}{b}\right)t}] = e^{\frac{a}{b}t} \cdot M_X\left(\dfrac{t}{b}\right)$. If $X_1, X_2, \ldots$, and X_n are independent random variables and $Y = X_1 + X_2 + \cdots + X_n$, then $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$ where $M_{X_i}(t)$ is the value of the moment-generating function of X_i at t", we can write

$$M_{\overline{X}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

and since the moment-generating function of a normal distribution with the mean μ and the variance σ^2 is given by

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

according to the theorem $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$, we get

$$M_{\overline{X}}(t) = \left[e^{\mu \cdot \frac{t}{n} + \frac{1}{2}\left(\frac{t}{n}\right)^2 \sigma^2}\right]^n$$

$$= e^{\mu t + \frac{1}{2}t^2\left(\frac{\sigma^2}{n}\right)}$$

This moment-generating function is readily seen to be that of a normal distribution with the mean μ and the variance σ^2/n.

3 The Sampling Distribution of the Mean: Finite Populations

If an experiment consists of selecting one or more values from a finite set of numbers $\{c_1, c_2, \ldots, c_N\}$, this set is referred to as a **finite population of size N**. In the definition that follows, it will be assumed that we are sampling *without replacement* from a finite population of size N.

DEFINITION 4. RANDOM SAMPLE—FINITE POPULATION. *If* X_1 *is the first value drawn from a finite population of size* N, X_2 *is the second value drawn,* $\ldots, X_n$ *is the* n*th value drawn, and the joint probability distribution of these* n *random variables is given by*

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{N(N-1) \cdot \ldots \cdot (N-n+1)}$$

for each ordered n*-tuple of values of these random variables, then* $X_1, X_2, \ldots, X_n$ *are said to constitute a* ***random sample*** *from the given finite population.*

As in Definition 2, the random sample is a set of random variables, but here again it is common practice also to apply the term "random sample" to the values of the random variables, that is, to the actual numbers drawn.

From the joint probability distribution of Definition 4, it follows that the probability for each subset of n of the N elements of the finite population (regardless of the order in which the values are drawn) is

$$\frac{n!}{N(N-1) \cdot \ldots \cdot (N-n+1)} = \frac{1}{\binom{N}{n}}$$

This is often given as an alternative definition or as a criterion for the selection of a random sample of size n from a finite population of size N: *Each of the* $\binom{N}{n}$ *possible samples must have the same probability*.

It also follows from the joint probability distribution of Definition 4 that the marginal distribution of X_r is given by

$$f(x_r) = \frac{1}{N} \qquad \text{for } x_r = c_1, c_2, \ldots, c_N$$

for $r = 1, 2, \ldots, n$, and we refer to the mean and the variance of this discrete uniform distribution as the mean and the variance of the finite population. Therefore,

DEFINITION 5. SAMPLE MEAN AND VARIANCE—FINITE POPULATION. *The* ***sample mean*** *and the* ***sample variance*** *of the finite population* $\{c_1, c_2, \ldots, c_N\}$ *are*

$$\mu = \sum_{i=1}^{N} c_i \cdot \frac{1}{N} \quad \textit{and} \quad \sigma^2 = \sum_{i=1}^{N} (c_i - \mu)^2 \cdot \frac{1}{N}$$

Finally, it follows from the joint probability distribution of Definition 4 that the joint marginal distribution of any two of the random variables $X_1, X_2, \ldots, X_n$ is given by

$$g(x_r, x_s) = \frac{1}{N(N-1)}$$

for each ordered pair of elements of the finite population. Thus, we can prove the following theorem.

THEOREM 5. If X_r and X_s are the rth and sth random variables of a random sample of size n drawn from the finite population $\{c_1, c_2, \ldots, c_N\}$, then

$$\text{cov}(X_r, X_s) = -\frac{\sigma^2}{N-1}$$

Proof According to the definition given here "**COVARIANCE.** $\mu_{1,1}$ *is called the* ***covariance*** *of* X *and* Y, *and it is denoted by* σ_{XY}, *cov*(X, Y), or C(X, Y)",

$$\text{cov}(X_r, X_s) = \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \frac{1}{N(N-1)}(c_i - \mu)(c_j - \mu)$$

$$= \frac{1}{N(N-1)} \cdot \sum_{i=1}^{N}(c_i - \mu)\left[\sum_{\substack{j=1 \\ j \neq i}}^{N}(c_j - \mu)\right]$$

and since $\sum_{\substack{j=1 \\ j \neq i}}^{N}(c_j - \mu) = \sum_{j=1}^{N}(c_j - \mu) - (c_i - \mu) = -(c_i - \mu)$, we get

$$\text{cov}(X_r, X_s) = -\frac{1}{N(N-1)} \cdot \sum_{i=1}^{N}(c_i - \mu)^2$$

$$= -\frac{1}{N-1} \cdot \sigma^2$$

Making use of all these results, let us now prove the following theorem, which, for random samples from finite populations, corresponds to Theorem 1.

THEOREM 6. If $\overline{X}$ is the mean of a random sample of size n taken without replacement from a finite population of size N with the mean μ and the variance σ^2, then

$$E(\overline{X}) = \mu \quad \text{and} \quad \text{var}(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Proof Substituting $a_i = \frac{1}{N}$, $\text{var}(X_i) = \sigma^2$, and $\text{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}$ into the formula $E(Y) = \sum_{i=1}^{n} a_i E(X_i)$, we get

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n} \cdot \mu = \mu$$

and

$$\text{var}(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n^2} \cdot \sigma^2 + 2 \cdot \sum_{i<j}\sum \frac{1}{n^2}\left(-\frac{\sigma^2}{N-1}\right)$$

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

$$= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

It is of interest to note that the formulas we obtained for $\text{var}(\overline{X})$ in Theorems 1 and 6 differ only by the **finite population correction factor** $\frac{N-n}{N-1}$.† Indeed, when N is large compared to n, the difference between the two formulas for $\text{var}(\overline{X})$ is usually negligible, and the formula $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ is often used as an approximation when we are sampling from a large finite population. A general rule of thumb is to use this approximation when the sample does not constitute more than 5 percent of the population.

Exercises

1. This question has been intentionally omitted for this edition.

2. This question has been intentionally omitted for this edition.

3. With reference to Exercise 2, show that if the two samples come from normal populations, then $\overline{X}_1 - \overline{X}_2$ is a random variable having a normal distribution with the mean $\mu_1 - \mu_2$ and the variance $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. (*Hint*: Proceed as in the proof of Theorem 4.)

4. If $X_1, X_2, \ldots, X_n$ are independent random variables having identical Bernoulli distributions with the parameter θ, then $\overline{X}$ is the proportion of successes in n trials, which we denote by $\hat{\Theta}$. Verify that

(a) $E(\hat{\Theta}) = \theta$;

(b) $\text{var}(\hat{\Theta}) = \frac{\theta(1-\theta)}{n}$.

5. If the first n_1 random variables of Exercise 2 have Bernoulli distributions with the parameter θ_1 and the other n_2 random variables have Bernoulli distributions with the parameter θ_2, show that, in the notation of Exercise 4,

(a) $E(\hat{\Theta}_1 - \hat{\Theta}_2) = \theta_1 - \theta_2$;

(b) $\text{var}(\hat{\Theta}_1 - \hat{\Theta}_2) = \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$.

6. This question has been intentionally omitted for this edition.

7. The following is a sufficient condition for the central limit theorem: If the random variables $X_1, X_2, \ldots, X_n$ are independent and uniformly bounded (that is, there exists a positive constant k such that the probability is zero that any one of the random variables X_i will take on a value greater than k or less than $-k$), then if the variance of

$$Y_n = X_1 + X_2 + \cdots + X_n$$

becomes infinite when $n \to \infty$, the distribution of the standardized mean of the X_i approaches the standard normal distribution. Show that this sufficient condition holds for a sequence of independent random variables X_i having the respective probability distributions

$$f_i(x_i) = \begin{cases} \frac{1}{2} & \text{for } x_i = 1 - (\frac{1}{2})^i \\ \frac{1}{2} & \text{for } x_i = (\frac{1}{2})^i - 1 \end{cases}$$

8. Consider the sequence of independent random variables $X_1, X_2, X_3, \ldots$ having the uniform densities

$$f_i(x_i) = \begin{cases} \frac{1}{2 - \frac{1}{i}} & \text{for } 0 < x_i < 2 - \frac{1}{i} \\ 0 & \text{elsewhere} \end{cases}$$

Use the sufficient condition of Exercise 7 to show that the central limit theorem holds.

9. The following is a sufficient condition, the *Laplace–Liapounoff condition*, for the central limit theorem: If $X_1, X_2, X_3, \ldots$ is a sequence of independent random variables, each having an absolute third moment

$$c_i = E(|X_i - \mu_i|^3)$$

and if

$$\lim_{n \to \infty} [\text{var}(Y_n)]^{-\frac{3}{2}} \cdot \sum_{i=1}^{n} c_i = 0$$

where $Y_n = X_1 + X_2 + \cdots + X_n$, then the distribution of the standardized mean of the X_i approaches the standard normal distribution when $n \to \infty$. Use this condition to show that the central limit theorem holds for the sequence of random variables of Exercise 7.

10. Use the condition of Exercise 9 to show that the central limit theorem holds for the sequence of random variables of Exercise 8.

†Since there are many problems in which we are interested in the standard deviation rather than the variance, the term "finite population correction factor" often refers to $\sqrt{\frac{N-n}{N-1}}$ instead of $\frac{N-n}{N-1}$. This does not matter, of course, as long as the usage is clearly understood.

11. Explain why, when we sample with replacement from a finite population, the results of Theorem 1 apply rather than those of Theorem 6.

12. This question has been intentionally omitted for this edition.

13. Use MINITAB or some other statistical computer program to generate 20 samples of size 10 each from the uniform density function $f(x) = 1, 0 \le x \le 1$.
(a) Find the mean of each sample and construct a histogram of these sample means.
(b) Calculate the mean and the variance of the 20 sample means.

14. Referring to Exercise 13, now change the sample size to 30.
(a) Does this histogram more closely resemble that of a normal distribution than that of Exercise 13? Why?
(b) Which resembles it more closely?
(c) Calculate the mean and the variance of the 20 sample means.

15. If a random sample of size n is selected without replacement from the finite population that consists of the integers $1, 2, \ldots, N$, show that
(a) the mean of $\overline{X}$ is $\dfrac{N+1}{2}$;
(b) the variance of $\overline{X}$ is $\dfrac{(N+1)(N-n)}{12n}$;
(c) the mean and the variance of $Y = n \cdot \overline{X}$ are

$$E(Y) = \frac{n(N+1)}{2} \quad \text{and} \quad \text{var}(Y) = \frac{n(N+1)(N-n)}{12}$$

16. Find the mean and the variance of the finite population that consists of the 10 numbers 15, 13, 18, 10, 6, 21, 7, 11, 20, and 9.

17. Show that the variance of the finite population $\{c_1, c_2, \ldots, c_N\}$ can be written as

$$\sigma^2 = \frac{\sum_{i=1}^{N} c_i^2}{N} - \mu^2$$

Also, use this formula to recalculate the variance of the finite population of Exercise 16.

18. Show that, analogous to the formula of Exercise 17, the formula for the sample variance can be written as

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{n\overline{X}^2}{n-1}$$

Also, use this formula to calculate the variance of the following sample data on the number of service calls received by a tow truck operator on eight consecutive working days: 13, 14, 13, 11, 15, 14, 17, and 11.

19. Show that the formula for the sample variance can be written as

$$S^2 = \frac{n\left(\sum_{i=1}^{n} X_i^2\right) - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

Also, use this formula to recalculate the variance of the sample data of Exercise 18.

4 The Chi-Square Distribution

If X has the standard normal distribution, then X^2 has the special gamma distribution, which is referred to as the **chi-square distribution**, and this accounts for the important role that the chi-square distribution plays in problems of sampling from normal populations. Theorem 11 will show the importance of this distribution in making inferences about sample variances.

The chi-square distribution is often denoted by "χ^2 distribution," where χ is the lowercase Greek letter *chi*. We also use χ^2 for values of random variables having chi-square distributions, but we shall refrain from denoting the corresponding random variables by X^2, where X is the capital Greek letter *chi*. This avoids having to reiterate in each case whether X is a random variable with values x or a random variable with values χ.

If a random variable X has the chi-square distribution with ν degrees of freedom if its probability density is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The mean and the variance of the chi-square distribution with ν degrees of freedom are ν and 2ν, respectively, and its moment-generating function is given by

$$M_X(t) = (1-2t)^{-\nu/2}$$

The chi-square distribution has several important mathematical properties, which are given in Theorems 7 through 10.

THEOREM 7. If X has the standard normal distribution, then X^2 has the chi-square distribution with $\nu = 1$ degree of freedom.

More generally, let us prove the following theorem.

THEOREM 8. If $X_1, X_2, \ldots, X_n$ are independent random variables having standard normal distributions, then

$$Y = \sum_{i=1}^{n} X_i^2$$

has the chi-square distribution with $\nu = n$ degrees of freedom.

Proof Using the moment-generating function given previously with $\nu = 1$ and Theorem 7, we find that

$$M_{X_i^2}(t) = (1-2t)^{-\frac{1}{2}}$$

and it follows the theorem "$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$" that

$$M_Y(t) = \prod_{i=1}^{n}(1-2t)^{-\frac{1}{2}} = (1-2t)^{-\frac{n}{2}}$$

This moment-generating function is readily identified as that of the chi-square distribution with $\nu = n$ degrees of freedom.

Two further properties of the chi-square distribution are given in the two theorems that follow; the reader will be asked to prove them in Exercises 20 and 21.

THEOREM 9. If $X_1, X_2, \ldots, X_n$ are independent random variables having chi-square distributions with $\nu_1, \nu_2, \ldots, \nu_n$ degrees of freedom, then

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

has the chi-square distribution with $\nu_1 + \nu_2 + \cdots + \nu_n$ degrees of freedom.

THEOREM 10. If X_1 and X_2 are independent random variables, X_1 has a chi-square distribution with ν_1 degrees of freedom, and $X_1 + X_2$ has a chi-square distribution with $\nu > \nu_1$ degrees of freedom, then X_2 has a chi-square distribution with $\nu - \nu_1$ degrees of freedom.

The chi-square distribution has many important applications. Foremost are those based, directly or indirectly, on the following theorem.

THEOREM 11. If $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the standard deviation σ, then

1. $\overline{X}$ and S^2 are independent;
2. the random variable $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $n - 1$ degrees of freedom.

Proof Since a detailed proof of part 1 would go beyond the scope of this chapter we shall assume the independence of $\overline{X}$ and S^2 in our proof of part 2. In addition to the references to proofs of part 1 at the end of this chapter, Exercise 31 outlines the major steps of a somewhat simpler proof based on the idea of a conditional moment-generating function, and in Exercise 30 the reader will be asked to prove the independence of $\overline{X}$ and S^2 for the special case where $n = 2$.

To prove part 2, we begin with the identity

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

which the reader will be asked to verify in Exercise 22. Now, if we divide each term by σ^2 and substitute $(n-1)S^2$ for $\sum_{i=1}^{n}(X_i-\overline{X})^2$, it follows that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

With regard to the three terms of this identity, we know from Theorem 8 that the one on the left-hand side of the equation is a random variable having a chi-square distribution with n degrees of freedom. Also, according to Theorems 4 and 7, the second term on the right-hand side of the equation is a random variable having a chi-square distribution with 1 degree of freedom. Now, since $\overline{X}$ and S^2 are assumed to be independent, it follows that the two terms on the right-hand side of the equation are independent, and we conclude that $\dfrac{(n-1)S^2}{\sigma^2}$ is a random variable having a chi-square distribution with $n-1$ degrees of freedom.

Since the chi-square distribution arises in many important applications, integrals of its density have been extensively tabulated. Table V of "Statistical Tables" contains values of $\chi^2_{\alpha,\nu}$ for $\alpha = 0.995, 0.99, 0.975, 0.95, 0.05, 0.025, 0.01, 0.005$, and $\nu = 1, 2, \ldots, 30$, where $\chi^2_{\alpha,\nu}$ is such that the area to its right under the chi-square curve with ν degrees of freedom (see Figure 1) is equal to α. That is, $\chi^2_{\alpha,\nu}$ is such that if X is a random variable having a chi-square distribution with ν degrees of freedom, then

$$P(X \geqq \chi^2_{\alpha,\nu}) = \alpha$$

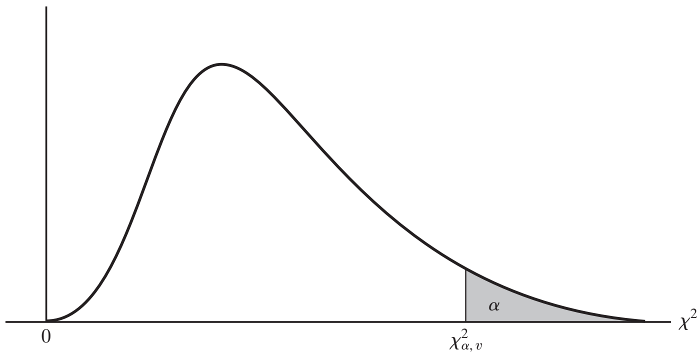

Figure 1. Chi-square distribution.

When ν is greater than 30, Table V of "Statistical Tables" cannot be used and probabilities related to chi-square distributions are usually approximated with normal distributions, as in Exercise 25 or 28.

EXAMPLE 2

Suppose that the thickness of a part used in a semiconductor is its critical dimension and that the process of manufacturing these parts is considered to be under control if the true variation among the thicknesses of the parts is given by a standard deviation not greater than $\sigma = 0.60$ thousandth of an inch. To keep a check on the process, random samples of size $n = 20$ are taken periodically, and it is regarded to be "out of control" if the probability that S^2 will take on a value greater than or equal to the observed sample value is 0.01 or less (even though $\sigma = 0.60$). What can one conclude about the process if the standard deviation of such a periodic random sample is $s = 0.84$ thousandth of an inch?

Solution

The process will be declared "out of control" if $\dfrac{(n-1)s^2}{\sigma^2}$ with $n = 20$ and $\sigma = 0.60$ exceeds $\chi^2_{0.01,19} = 36.191$. Since

$$\frac{(n-1)s^2}{\sigma^2} = \frac{19(0.84)^2}{(0.60)^2} = 37.24$$

exceeds 36.191, the process is declared out of control. Of course, it is assumed here that the sample may be regarded as a random sample from a normal population.

---

5 The *t* Distribution

In Theorem 4 we showed that for random samples from a normal population with the mean μ and the variance σ^2, the random variable $\overline{X}$ has a normal distribution with the mean μ and the variance $\dfrac{\sigma^2}{n}$; in other words,

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

has the standard normal distribution. This is an important result, but the major difficulty in applying it is that in most realistic applications the population standard deviation σ is unknown. This makes it necessary to replace σ with an estimate, usually with the value of the sample standard deviation S. Thus, the theory that follows leads to the exact distribution of $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ for random samples from normal populations.

To derive this sampling distribution, let us first study the more general situation treated in the following theorem.

Theorem 12. If Y and Z are independent random variables, Y has a chi-square distribution with ν degrees of freedom, and Z has the standard normal distribution, then the distribution of

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{for } -\infty < t < \infty$$

and it is called the ***t* distribution with *ν* degrees of freedom**.

Proof Since Y and Z are independent, their joint probability density is given by

$$f(y,z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \cdot \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}}y^{\frac{\nu}{2}-1}e^{-\frac{y}{2}}$$

for $y > 0$ and $-\infty < z < \infty$, and $f(y,z) = 0$ elsewhere. Then, to use the change-of-variable technique, we solve $t = \dfrac{z}{\sqrt{y/\nu}}$ for z, getting $z = t\sqrt{y/\nu}$ and hence $\dfrac{\partial z}{\partial t} = \sqrt{y/\nu}$. Thus, the joint density of Y and T is given by

$$g(y,t) = \begin{cases} \dfrac{1}{\sqrt{2\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}}y^{\frac{\nu-1}{2}}e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)} & \text{for } y > 0 \text{ and } -\infty < t < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and, integrating out y with the aid of the substitution $w = \dfrac{y}{2}\left(1+\dfrac{t^2}{\nu}\right)$, we finally get

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{for } -\infty < t < \infty$$

The t distribution was introduced originally by W. S. Gosset, who published his scientific writings under the pen name "Student," since the company for which he worked, a brewery, did not permit publication by employees. Thus, the t distribution is also known as the **Student t distribution**, or **Student's t distribution**. As shown in Figure 2, graphs of t distributions having different numbers of degrees of freedom resemble that of the standard normal distribution, but have larger variances. In fact, for large values of υ, the t distribution approaches the standard normal distribution.

In view of its importance, the t distribution has been tabulated extensively. Table IV of "Statistical Tables", for example, contains values of $t_{\alpha,\nu}$ for $\alpha = 0.10, 0.05, 0.025, 0.01, 0.005$ and $\nu = 1, 2, \ldots, 29$, where $t_{\alpha,\nu}$ is such that the area to its right under the curve of the t distribution with ν degrees of freedom (see Figure 3) is equal to α. That is, $t_{\alpha,\nu}$ is such that if T is a random variable having a t distribution with ν degrees of freedom, then

$$P(T \geqq t_{\alpha,\nu}) = \alpha$$

The table does not contain values of $t_{\alpha,\nu}$ for $\alpha > 0.50$, since the density is symmetrical about $t = 0$ and hence $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$. When ν is 30 or more, probabilities related to the t distribution are usually approximated with the use of normal distributions (see Exercise 35).

Among the many applications of the t distribution, its major application (for which it was originally developed) is based on the following theorem.

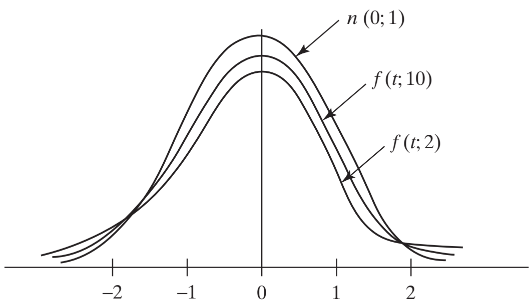

Figure 2. Comparison of t distributions and standard normal distribution.

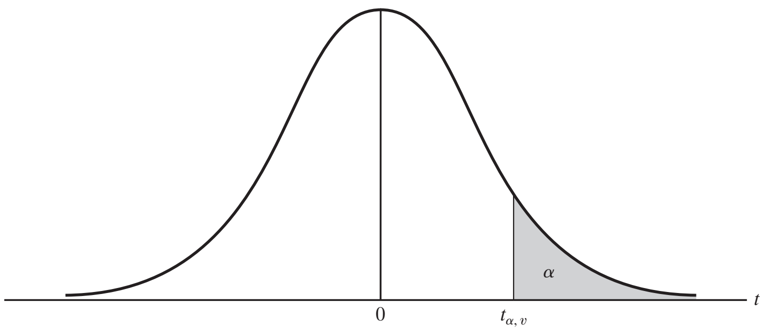

Figure 3. t distribution.

Theorem 13. If $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the variance σ^2, then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has the t distribution with $n - 1$ degrees of freedom.

Proof By Theorems 11 and 4, the random variables

$$Y = \frac{(n-1)S^2}{\sigma^2} \quad \text{and} \quad Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

have, respectively, a chi-square distribution with $n - 1$ degrees of freedom and the standard normal distribution. Since they are also independent by part 1 of Theorem 11, substitution into the formula for T of Theorem 12 yields

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

and this completes the proof.

EXAMPLE 3

In 16 one-hour test runs, the gasoline consumption of an engine averaged 16.4 gallons with a standard deviation of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour.

Solution

Substituting $n = 16, \mu = 12.0$, $\bar{x} = 16.4$, and $s = 2.1$ into the formula for t in Theorem 13, we get

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.4 - 12.0}{2.1/\sqrt{16}} = 8.38$$

Since Table IV of "Statistical Tables" shows that for $\nu = 15$ the probability of getting a value of T greater than 2.947 is 0.005, the probability of getting a value greater than 8 must be negligible. Thus, it would seem reasonable to conclude that the true average hourly gasoline consumption of the engine exceeds 12.0 gallons.

6 The *F* Distribution

Another distribution that plays an important role in connection with sampling from normal populations is the F distribution, named after Sir Ronald A. Fisher, one of the most prominent statisticians of the last century. Originally, it was studied as the sampling distribution of the ratio of two independent random variables with chi-square distributions, each divided by its respective degrees of freedom, and this is how we shall present it here.

Fisher's F distribution is used to draw statistical inferences about the ratio of two sample variances. As such, it plays a key role in the analysis of variance, used in conjunction with experimental designs.

THEOREM 14. If U and V are independent random variables having chi-square distributions with ν_1 and ν_2 degrees of freedom, then

$$F = \frac{U/\nu_1}{V/\nu_2}$$

is a random variable having an ***F* distribution**, that is, a random variable whose probability density is given by

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

for $f > 0$ and $g(f) = 0$ elsewhere.

Proof By virtue of independence, the joint density of U and V is given by

$$\begin{aligned} f(u,v) &= \frac{1}{2^{\nu_1/2}\Gamma\left(\frac{\nu_1}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1}e^{-\frac{u}{2}} \cdot \frac{1}{2^{\nu_2/2}\Gamma\left(\frac{\nu_2}{2}\right)} \cdot v^{\frac{\nu_2}{2}-1}e^{-\frac{v}{2}} \\ &= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1}v^{\frac{\nu_2}{2}-1}e^{-\frac{u+v}{2}} \end{aligned}$$

for $u > 0$ and $v > 0$, and $f(u,v) = 0$ elsewhere. Then, to use the change-of-variable technique, we solve

$$f = \frac{u/\nu_1}{v/\nu_2}$$

for u, getting $u = \frac{\nu_1}{\nu_2} \cdot vf$ and hence $\frac{\partial u}{\partial f} = \frac{\nu_1}{\nu_2} \cdot v$. Thus, the joint density of F and V is given by

$$g(f,v) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot f^{\frac{\nu_1}{2}-1}v^{\frac{\nu_1+\nu_2}{2}-1}e^{-\frac{v}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)}$$

for $f > 0$ and $v > 0$, and $g(f,v) = 0$ elsewhere. Now, integrating out v by making the substitution $w = \frac{v}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)$, we finally get

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

for $f > 0$, and $g(f) = 0$ elsewhere.

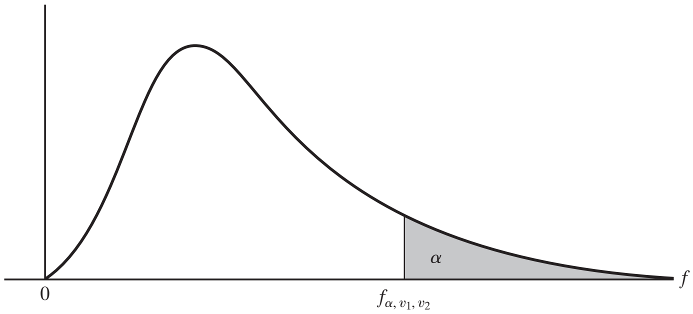

Figure 4. *F* distribution.

In view of its importance, the F distribution has been tabulated extensively. Table VI of "Statistical Tables", for example, contains values of f_{α,ν_1,ν_2} for $\alpha = 0.05$ and 0.01 and for various values of ν_1 and ν_2, where f_{α,ν_1,ν_2} is such that the area to its right under the curve of the F distribution with ν_1 and ν_2 degrees of freedom (see Figure 4) is equal to α. That is, f_{α,ν_1,ν_2} is such that

$$P(F \geqq f_{\alpha,\nu_1,\nu_2}) = \alpha$$

Applications of Theorem 14 arise in problems in which we are interested in comparing the variances σ_1^2 and σ_2^2 of two normal populations; for instance, in problems in which we want to estimate the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ or perhaps to test whether $\sigma_1^2 = \sigma_2^2$. We base such inferences on **independent random samples** of sizes n_1 and n_2 from the two populations and Theorem 11, according to which

$$\chi_1^2 = \frac{(n_1 - 1)s_1^2}{\sigma_1^2} \quad \text{and} \quad \chi_2^2 = \frac{(n_2 - 1)s_2^2}{\sigma_2^2}$$

are values of random variables having chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. By "independent random samples," we mean that the $n_1 + n_2$ random variables constituting the two random samples are all independent, so that the two chi-square random variables are independent and the substitution of their values for U and V in Theorem 14 yields the following result.

THEOREM 15. If S_1^2 and S_2^2 are the variances of independent random samples of sizes n_1 and n_2 from normal populations with the variances σ_1^2 and σ_2^2, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

is a random variable having an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

The F distribution is also known as the **variance-ratio distribution**.

Exercises

20. Prove Theorem 9.

21. Prove Theorem 10.

22. Verify the identity

$$\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2$$

which we used in the proof of Theorem 11.

23. Use Theorem 11 to show that, for random samples of size n from a normal population with the variance σ^2, the sampling distribution of S^2 has the mean σ^2 and the variance $\dfrac{2\sigma^4}{n-1}$. (A general formula for the variance of S^2 for random samples from any population with finite second and fourth moments may be found in the book by H. Cramér listed among the references at the end of this chapter.)

24. Show that if $X_1, X_2, \ldots, X_n$ are independent random variables having the chi-square distribution with $\nu = 1$ and $Y_n = X_1+X_2+\cdots+X_n$, then the limiting distribution of

$$Z=\frac{\frac{Y_n}{n}-1}{\sqrt{2/n}}$$

as $n\to\infty$ is the standard normal distribution.

25. Based on the result of Exercise 24, show that if X is a random variable having a chi-square distribution with ν degrees of freedom and ν is large, the distribution of $\dfrac{X-\nu}{\sqrt{2\nu}}$ can be approximated with the standard normal distribution.

26. Use the method of Exercise 25 to find the approximate value of the probability that a random variable having a chi-square distribution with $\nu = 50$ will take on a value greater than 68.0.

27. If the range of X is the set of all positive real numbers, show that for $k > 0$ the probability that $\sqrt{2X}-\sqrt{2\nu}$ will take on a value less than k equals the probability that $\dfrac{X-\nu}{\sqrt{2\nu}}$ will take on a value less than $k+\dfrac{k^2}{2\sqrt{2\nu}}$.

28. Use the results of Exercises 25 and 27 to show that if X has a chi-square distribution with ν degrees of freedom, then for large ν the distribution of $\sqrt{2X}-\sqrt{2\nu}$ can be approximated with the standard normal distribution. Also, use this method of approximation to rework Exercise 26.

29. Find the percentage errors of the approximations of Exercises 26 and 28, given that the actual value of the probability (rounded to five decimals) is 0.04596.

30. (*Proof of the independence of* $\overline{X}$ *and* S^2 *for* $n = 2$) If X_1 and X_2 are independent random variables having the standard normal distribution, show that

(a) the joint density of X_1 and $\overline{X}$ is given by

$$f(x_1,\bar{x})=\frac{1}{\pi}\cdot e^{-\bar{x}^{2}}e^{-(x_1-\bar{x})^2}$$

for $-\infty < x_1 < \infty$ and $-\infty < \bar{x} < \infty$;

(b) the joint density of $U = |X_1-\overline{X}|$ and $\overline{X}$ is given by

$$g(u,\bar{x})=\frac{2}{\pi}\cdot e^{-(\bar{x}^2+u^2)}$$

for $u > 0$ and $-\infty < \bar{x} < \infty$, since $f(x_1,\bar{x})$ is symmetrical about $\bar{x}$ for fixed $\bar{x}$;

(c) $S^2 = 2(X_1-\overline{X})^2 = 2U^2$;

(d) the joint density of $\overline{X}$ and S^2 is given by

$$h(s^2,\bar{x})=\frac{1}{\sqrt{\pi}}e^{-\bar{x}^2}\cdot\frac{1}{\sqrt{2\pi}}(s^2)^{-\frac{1}{2}}e^{-\frac{1}{2}s^2}$$

for $s^2 > 0$ and $-\infty < \bar{x} < \infty$, demonstrating that $\overline{X}$ and S^2 are independent.

31. (*Proof of the independence of* $\overline{X}$ *and* S^2) If $X_1, X_2, \ldots, X_n$ constitute a random sample from a normal population with the mean μ and the variance σ^2,

(a) find the conditional density of X_1 given $X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n$, and then set $X_1 = n\overline{X}-X_2-\cdots-X_n$ and use the transformation technique to find the conditional density of $\overline{X}$ given $X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n$;

(b) find the joint density of $\overline{X}, X_2, X_3, \ldots, X_n$ by multiplying the conditional density of $\overline{X}$ obtained in part (a) by the joint density of $X_2, X_3, \ldots, X_n$, and show that

$$g(x_2,x_3,\ldots,x_n|\bar{x})=\sqrt{n}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1}e^{-\frac{(n-1)s^2}{2\sigma^2}}$$

for $-\infty < x_i < \infty, i = 2, 3, \ldots, n$;

(c) show that the conditional moment-generating function of $\dfrac{(n-1)S^2}{\sigma^2}$ given $\overline{X}=\bar{x}$ is

$$E\left[e^{\frac{(n-1)S^2}{\sigma^2}\cdot t}\middle|\bar{x}\right]=(1-2t)^{-\frac{n-1}{2}}\qquad\text{for } t<\frac{1}{2}$$

Since this result is free of $\bar{x}$, it follows that $\overline{X}$ and S^2 are independent; it also shows that $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $n-1$ degrees of freedom.

This proof, due to J. Shuster, is listed among the references at the end of this chapter.

32. This question has been intentionally omitted for this edition.

33. Show that for $\nu > 2$ the variance of the t distribution with ν degrees of freedom is $\dfrac{\nu}{\nu-2}$. (*Hint*: Make the substitution $1+\dfrac{t^2}{\nu}=\dfrac{1}{u}$.)

34. Show that for the t distribution with $\nu > 4$ degrees of freedom

(a) $\mu_4 = \dfrac{3\nu^2}{(\nu-2)(\nu-4)}$;

(b) $\alpha_4 = 3+\dfrac{6}{\nu-4}$.

(*Hint*: Make the substitution $1+\dfrac{t^2}{\nu}=\dfrac{1}{u}$.)

35. This question has been intentionally omitted for this edition.

36. By what name did we refer to the t distribution with $\nu = 1$ degree of freedom?

37. This question has been intentionally omitted for this edition.

38. Show that for $\nu_2 > 2$ the mean of the F distribution is $\dfrac{\nu_2}{\nu_2-2}$, making use of the definition of F in Theorem 14 and the fact that for a random variable V having the chi-square distribution with ν_2 degrees of freedom, $E\left(\dfrac{1}{V}\right)=\dfrac{1}{\nu_2-2}$.

39. Verify that if X has an F distribution with ν_1 and ν_2 degrees of freedom and $\nu_2 \to \infty$, the distribution of $Y = \nu_1 X$ approaches the chi-square distribution with ν_1 degrees of freedom.

40. Verify that if T has a t distribution with ν degrees of freedom, then $X = T^2$ has an F distribution with $\nu_1 = 1$ and $\nu_2 = \nu$ degrees of freedom.

41. If X has an F distribution with ν_1 and ν_2 degrees of freedom, show that $Y = \dfrac{1}{X}$ has an F distribution with ν_2 and ν_1 degrees of freedom.

42. Use the result of Exercise 41 to show that

$$f_{1-\alpha,\nu_1,\nu_2} = \frac{1}{f_{\alpha,\nu_2,\nu_1}}$$

43. Verify that if Y has a beta distribution with $\alpha = \dfrac{\nu_1}{2}$ and $\beta = \dfrac{\nu_2}{2}$, then

$$X = \frac{\nu_2 Y}{\nu_1(1-Y)}$$

has an F distribution with ν_1 and ν_2 degrees of freedom.

44. Show that the F distribution with 4 and 4 degrees of freedom is given by

$$g(f) = \begin{cases} 6f(1+f)^{-4} & \text{for } f > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and use this density to find the probability that for independent random samples of size $n = 5$ from normal populations with the same variance, S_1^2/S_2^2 will take on a value less than $\frac{1}{2}$ or greater than 2.

7 Order Statistics

The sampling distributions presented so far in this chapter depend on the assumption that the population from which the sample was taken has the normal distribution. This assumption often is satisfied, at least approximately for large samples, as illustrated by the central limit theorem. However, small samples sometimes must be used in practice, for example in statistical quality control or where taking and measuring a sample is very expensive. In an effort to deal with the problem of small samples in cases where it may be unreasonable to assume a normal population, statisticians have developed **nonparametric statistics**, whose sampling distributions do not depend upon any assumptions about the population from which the sample is taken. Statistical inferences based upon such statistics are called **nonparametric** inference. We will identify a class of nonparametric statistics called **order statistics** and discuss their statistical properties.

Consider a random sample of size n from an infinite population with a continuous density, and suppose that we arrange the values of $X_1, X_2, \ldots$, and X_n according to size. If we look upon the smallest of the x's as a value of the random variable Y_1, the next largest as a value of the random variable Y_2, the next largest after that as a value of the random variable $Y_3, \ldots$, and the largest as a value of the random variable Y_n, we refer to these Y's as **order statistics**. In particular, Y_1 is the first order statistic, Y_2 is the second order statistic, Y_3 is the third order statistic, and so on. (We are limiting this discussion to infinite populations with continuous densities so that there is zero probability that any two of the x's will be alike.)

To be more explicit, consider the case where $n = 2$ and the relationship between the values of the X's and the Y's is

$$\begin{array}{llll} y_1 = x_1 & \text{and} \quad y_2 = x_2 & \text{when} & x_1 < x_2 \\ y_1 = x_2 & \text{and} \quad y_2 = x_1 & \text{when} & x_2 < x_1 \end{array}$$

Similarly, for $n = 3$ the relationship between the values of the respective random variables is

$$\begin{array}{llllll} y_1 = x_1, & y_2 = x_2, & \text{and} & y_3 = x_3, & \text{when} & x_1 < x_2 < x_3 \\ y_1 = x_1, & y_2 = x_3, & \text{and} & y_3 = x_2, & \text{when} & x_1 < x_3 < x_2 \\ \multicolumn{6}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ y_1 = x_3, & y_2 = x_2, & \text{and} & y_3 = x_1, & \text{when} & x_3 < x_2 < x_1 \end{array}$$

Let us now derive a formula for the probability density of the rth order statistic for $r = 1, 2, \ldots, n$.

THEOREM 16. For random samples of size n from an infinite population that has the value $f(x)$ at x, the probability density of the rth order statistic Y_r is given by

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}$$

for $-\infty < y_r < \infty$.

Proof Suppose that the real axis is divided into three intervals, one from $-\infty$ to y_r, a second from y_r to $y_r + h$ (where h is a positive constant), and the third from $y_r + h$ to ∞. Since the population we are sampling has the value $f(x)$ at x, the probability that $r - 1$ of the sample values fall into the first interval, 1 falls into the second interval, and $n - r$ fall into the third interval is

$$\frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1}\left[\int_{y_r}^{y_r+h} f(x)\,dx\right]\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r}$$

according to the formula for the multinomial distribution. Using the mean-value theorem for integrals from calculus, we have

$$\int_{y_r}^{y_r+h} f(x)\,dx = f(\xi)\cdot h \qquad \text{where } y_r \leqq \xi \leqq y_r + h$$

and if we let $h \to 0$, we finally get

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}$$

for $-\infty < y_r < \infty$ for the probability density of the rth order statistic.

In particular, the sampling distribution of Y_1, the smallest value in a random sample of size n, is given by

$$g_1(y_1) = n \cdot f(y_1)\left[\int_{y_1}^{\infty} f(x)\,dx\right]^{n-1} \qquad \text{for } -\infty < y_1 < \infty$$

while the sampling distribution of Y_n, the largest value in a random sample of size n, is given by

$$g_n(y_n) = n \cdot f(y_n)\left[\int_{-\infty}^{y_n} f(x)\,dx\right]^{n-1} \qquad \text{for } -\infty < y_n < \infty$$

Also, in a random sample of size $n = 2m+1$ the **sample median** $\tilde{X}$ is Y_{m+1}, whose sampling distribution is given by

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!}\left[\int_{-\infty}^{\tilde{x}} f(x)\,dx\right]^{m} f(\tilde{x})\left[\int_{\tilde{x}}^{\infty} f(x)\,dx\right]^{m} \qquad \text{for } -\infty < \tilde{x} < \infty$$

[For random samples of size $n = 2m$, the median is defined as $\frac{1}{2}(Y_m + Y_{m+1})$.]

In some instances it is possible to perform the integrations required to obtain the densities of the various order statistics; for other populations there may be no choice but to approximate these integrals by using numerical methods.

EXAMPLE 4

Show that for random samples of size n from an exponential population with the parameter θ, the sampling distributions of Y_1 and Y_n are given by

$$g_1(y_1) = \begin{cases} \dfrac{n}{\theta} \cdot e^{-ny_1/\theta} & \text{for } y_1 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$g_n(y_n) = \begin{cases} \dfrac{n}{\theta} \cdot e^{-y_n/\theta}[1 - e^{-y_n/\theta}]^{n-1} & \text{for } y_n > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and that, for random samples of size $n = 2m+1$ from this kind of population, the sampling distribution of the median is given by

$$h(\tilde{x}) = \begin{cases} \dfrac{(2m+1)!}{m!m!\theta} \cdot e^{-\tilde{x}(m+1)/\theta}[1 - e^{-\tilde{x}/\theta}]^{m} & \text{for } \tilde{x} > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Solution
The integrations required to obtain these results are straightforward, and they will be left to the reader in Exercise 45.

The following is an interesting result about the sampling distribution of the median, which holds when the population density is continuous and nonzero at the **population median** $\tilde{\mu}$, which is such that $\int_{-\infty}^{\tilde{\mu}} f(x)\,dx = \frac{1}{2}$.

> **THEOREM 17.** For large n, the sampling distribution of the median for random samples of size $2n+1$ is approximately normal with the mean $\tilde{\mu}$ and the variance $\dfrac{1}{8[f(\tilde{\mu})]^2 n}$.

Note that for random samples of size $2n+1$ from a normal population we have $\mu = \tilde{\mu}$, so

$$f(\tilde{\mu}) = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

and the variance of the median is approximately $\dfrac{\pi\sigma^2}{4n}$. If we compare this with the variance of the mean, which for random samples of size $2n+1$ from an infinite population is $\dfrac{\sigma^2}{2n+1}$, we find that for large samples from normal populations the mean is **more reliable** than the median; that is, the mean is subject to smaller chance fluctuations than the median.

Exercises

45. Verify the results of Example 4, that is, the sampling distributions of Y_1, Y_n, and $\tilde{X}$ shown there for random samples from an exponential population.

46. Find the sampling distributions of Y_1 and Y_n for random samples of size n from a continuous uniform population with $\alpha = 0$ and $\beta = 1$.

47. Find the sampling distribution of the median for random samples of size $2m+1$ from the population of Exercise 46.

48. Find the mean and the variance of the sampling distribution of Y_1 for random samples of size n from the population of Exercise 46.

49. Find the sampling distributions of Y_1 and Y_n for random samples of size n from a population having the beta distribution with $\alpha = 3$ and $\beta = 2$.

50. Find the sampling distribution of the median for random samples of size $2m+1$ from the population of Exercise 49.

51. Find the sampling distribution of Y_1 for random samples of size $n = 2$ taken
(a) without replacement from the finite population that consists of the first five positive integers;
(b) with replacement from the same population.
(*Hint*: Enumerate all possibilities.)

52. Duplicate the method used in the proof of Theorem 16 to show that the joint density of Y_1 and Y_n is given by

$$g(y_1, y_n) = n(n-1)f(y_1)f(y_n)\left[\int_{y_1}^{y_n} f(x)\,dx\right]^{n-2} \quad \text{for } -\infty < y_1 < y_n < \infty$$

and $g(y_1, y_n) = 0$ elsewhere.
(a) Use this result to find the joint density of Y_1 and Y_n for random samples of size n from an exponential population.
(b) Use this result to find the joint density of Y_1 and Y_n for the population of Exercise 46.

53. With reference to part (b) of Exercise 52, find the covariance of Y_1 and Y_n.

54. Use the formula for the joint density of Y_1 and Y_n shown in Exercise 52 and the transformation technique of several variables to find an expression for the joint density of Y_1 and the **sample range** $R = Y_n - Y_1$.

55. Use the result of Exercise 54 and that of part (a) of Exercise 52 to find the sampling distribution of R for random samples of size n from an exponential population.

56. Use the result of Exercise 54 to find the sampling distribution of R for random samples of size n from the continuous uniform population of Exercise 46.

57. Use the result of Exercise 56 to find the mean and the variance of the sampling distribution of R for random samples of size n from the continuous uniform population of Exercise 46.

58. There are many problems, particularly in industrial applications, in which we are interested in the proportion of a population that lies between certain limits. Such limits are called **tolerance limits**. The following steps lead to the sampling distribution of the statistic P, which is the proportion of a population (having a continuous density) that lies between the smallest and the largest values of a random sample of size n.

(a) Use the formula for the joint density of Y_1 and Y_n shown in Exercise 52 and the transformation technique of several variables to show that the joint density of Y_1 and P, whose values are given by

$$p = \int_{y_1}^{y_n} f(x)\,dx$$

is

$$h(y_1, p) = n(n-1)f(y_1)p^{n-2}$$

(b) Use the result of part (a) and the transformation technique of several variables to show that the joint density of P and W, whose values are given by

$$w = \int_{-\infty}^{y_1} f(x)\,dx$$

is

$$\varphi(w, p) = n(n-1)p^{n-2}$$

for $w > 0, p > 0, w + p < 1$, and $\varphi(w, p) = 0$ elsewhere.

(c) Use the result of part (b) to show that the marginal density of P is given by

$$g(p) = \begin{cases} n(n-1)p^{n-2}(1-p) & \text{for } 0 < p < 1 \\ 0 & \text{elsewhere} \end{cases}$$

This is the desired density of the proportion of the population that lies between the smallest and the largest values of a random sample of size n, and it is of interest to note that it does not depend on the form of the population distribution.

59. Use the result of Exercise 58 to show that, for the random variable P defined there,

$$E(P) = \frac{n-1}{n+1} \quad \text{and} \quad \text{var}(P) = \frac{2(n-1)}{(n+1)^2(n+2)}$$

What can we conclude from this about the distribution of P when n is large?

8 The Theory in Practice

More on Random Samples

While it is practically impossible to take a purely random sample, there are several methods that can be employed to assure that a sample is close enough to randomness to be useful in representing the distribution from which it came. In selecting a sample from a production line, *systematic sampling* can be used to select units at evenly spaced periods of time or having evenly spaced run numbers. In selecting a random sample from products in a warehouse, a *two-stage sampling process* can be used, numbering the containers and using a random device, such as a set of random numbers generated by a computer, to choose the containers. Then, a second set of random numbers can be used to select the unit or units in each container to be included in the sample. There are many other methods, employing mechanical devices or computer-generated random numbers, that can be used to aid in selecting a random sample.

Selection of a sample that reasonably can be regarded as random sometimes requires ingenuity, but it always requires care. Care should be taken to assure that only the specified distribution is represented. Thus, if a sample of product is meant to represent an entire production line, it should not be taken from the first shift only. Care should be taken to assure independence of the observations. Thus, the production-line sample should not be taken from a “chunk” of products produced at about the same time; they represent the same set of conditions and settings, and the resulting observations are closely related to each other. Human judgment in selecting samples usually includes personal bias, often unconscious, and such judgments should be avoided. Whenever possible, the use of mechanical devices or random numbers is preferable to methods involving personal choice.

The Assumption of Normality

It is not unusual to expect that errors are made in taking and recording observations. This phenomenon was described by early nineteenth-century astronomers who noted that different observers obtained somewhat different results when determining the location of a star.

Observational error can arise from one or both of two sources, **random error**, or statistical error, and **bias**. Random errors occur as the result of many imperfections of measurement; among these imperfections are imprecise markings on measurement scales, parallax (not viewing readings straight on) errors in setting up apparatus, slight differences in materials, expansion and contraction, minor changes in ambient conditions, and so forth. Bias occurs when there is a relatively consistent error, such as not obtaining a representative sample in a survey, using a measuring instrument that is not properly calibrated, and recording errors.

Errors involving bias can be corrected by discerning the source of the error and making appropriate “fixes” to eliminate the bias. Random error, however, is something we must live with, as no human endeavor can be made perfect in applications. Let us assume, however, that the many individual sources of random error, known or unknown, are additive. In fact this is usually the case, at least to a good approximation. Then we can write

$$X = \mu + E_1 + E_2 + \cdots + E_n$$

where the random variable X is an observed value, μ is the “true” value of the observation, and the E_i are the n random errors that affect the value of the observation. We shall assume that

$$E(X) = \mu + E(E_1) + E(E_2) + \cdots + E(E_n) = \mu$$

In other words, we are assuming that the random errors have a mean of zero, at least in the long run. We also can write

$$\text{var}(X) = (\mu + E_1 + E_2 + \cdots + E_n) = n\sigma^2$$

In other words, the variance of the *sum* of the random errors is $n\sigma^2$.

It follows that $\overline{X} = \mu + \overline{E}$, where $\overline{E}$ is the sample mean of the errors $E_1, E_2, \ldots,$ E_n, and $\sigma^2{}_{\overline{X}} = \sigma^2/n$. The central limit theorem given by Theorem 3 allows us to conclude that

$$Z = \frac{\overline{X} - \mu}{\sigma\sqrt{n}}$$

is a random variable whose distribution as $n \to \infty$ is the standard normal distribution.

It is not difficult to see from this argument that most repeated measurements of the same phenomenon are, at least approximately, normally distributed. It is this conclusion that underscores the importance of the chi-square, t, and F distributions in applications that are based on the assumption of data from normally distributed populations. It also demonstrates why the normal distribution is of major importance in statistics.

Applied Exercises SECS. 1–3

In the following exercises it is assumed that all samples are drawn without replacement unless otherwise specified.

60. How many different samples of size $n = 3$ can be drawn from a finite population of size

(a) $N = 12$; **(b)** $N = 20$; **(c)** $N = 50$?

61. What is the probability of each possible sample if

(a) a random sample of size $n = 4$ is to be drawn from a finite population of size $N = 12$;

(b) a random sample of size $n = 5$ is to be drawn from a finite population of size $N = 22$?

62. If a random sample of size $n = 3$ is drawn from a finite population of size $N = 50$, what is the probability that a particular element of the population will be included in the sample?

63. For random samples from an infinite population, what happens to the standard error of the mean if the sample size is

(a) increased from 30 to 120;

(b) increased from 80 to 180;

(c) decreased from 450 to 50;

(d) decreased from 250 to 40?

64. Find the value of the finite population correction factor $\frac{N-n}{N-1}$ for

(a) $n = 5$ and $N = 200$;

(b) $n = 50$ and $N = 300$;

(c) $n = 200$ and $N = 800$.

65. A random sample of size $n = 100$ is taken from an infinite population with the mean $\mu = 75$ and the variance $\sigma^2 = 256$.

(a) Based on Chebyshev's theorem, with what probability can we assert that the value we obtain for $\overline{X}$ will fall between 67 and 83?

(b) Based on the central limit theorem, with what probability can we assert that the value we obtain for $\overline{X}$ will fall between 67 and 83?

66. A random sample of size $n = 81$ is taken from an infinite population with the mean $\mu = 128$ and the standard deviation $\sigma = 6.3$. With what probability can we assert that the value we obtain for $\overline{X}$ will not fall between 126.6 and 129.4 if we use

(a) Chebyshev's theorem;

(b) the central limit theorem?

67. Rework part (b) of Exercise 66, assuming that the population is not infinite but finite and of size $N = 400$.

68. A random sample of size $n = 225$ is to be taken from an exponential population with $\theta = 4$. Based on the central limit theorem, what is the probability that the mean of the sample will exceed 4.5?

69. A random sample of size $n = 200$ is to be taken from a uniform population with $\alpha = 24$ and $\beta = 48$. Based on the central limit theorem, what is the probability that the mean of the sample will be less than 35?

70. A random sample of size 64 is taken from a normal population with $\mu = 51.4$ and $\sigma = 6.8$. What is the probability that the mean of the sample will

(a) exceed 52.9;

(b) fall between 50.5 and 52.3;

(c) be less than 50.6?

71. A random sample of size 100 is taken from a normal population with $\sigma = 25$. What is the probability that the mean of the sample will differ from the mean of the population by 3 or more either way?

72. Independent random samples of sizes 400 are taken from each of two populations having equal means and the standard deviations $\sigma_1 = 20$ and $\sigma_2 = 30$. Using Chebyshev's theorem and the result of Exercise 2, what can we assert with a probability of at least 0.99 about the value we will get for $\overline{X}_1 - \overline{X}_2$? (By "independent" we mean that the samples satisfy the conditions of Exercise 2.)

73. Assume that the two populations of Exercise 72 are normal and use the result of Exercise 3 to find k such that

$$P(-k < \overline{X}_1 - \overline{X}_2 < k) = 0.99$$

74. Independent random samples of sizes $n_1 = 30$ and $n_2 = 50$ are taken from two normal populations having the means $\mu_1 = 78$ and $\mu_2 = 75$ and the variances $\sigma_1^2 = 150$ and $\sigma_2^2 = 200$. Use the results of Exercise 3 to find the probability that the mean of the first sample will exceed that of the second sample by at least 4.8.

75. The actual proportion of families in a certain city who own, rather than rent, their home is 0.70. If 84 families in this city are interviewed at random and their responses to the question of whether they own their home are looked upon as values of independent random variables having identical Bernoulli distributions with the parameter $\theta = 0.70$, with what probability can we assert that the value we obtain for the sample proportion $\hat{\Theta}$ will fall between 0.64 and 0.76, using the result of Exercise 4 and

(a) Chebyshev's theorem;

(b) the central limit theorem?

76. The actual proportion of men who favor a certain tax proposal is 0.40 and the corresponding proportion for women is 0.25; $n_1 = 500$ men and $n_2 = 400$

women are interviewed at random, and their individual responses are looked upon as the values of independent random variables having Bernoulli distributions with the respective parameters $\theta_1 = 0.40$ and $\theta_2 = 0.25$. What can we assert, according to Chebyshev's theorem, with a probability of at least 0.9375 about the value we will get for $\hat{\Theta}_1 - \hat{\Theta}_2$, the difference between the two sample proportions of favorable responses? Use the result of Exercise 5.

SECS. 4–6

(In Exercises 78 through 83, refer to Tables IV, V, and VI of "Statistical Tables.")

77. Integrate the appropriate chi-square density to find the probability that the variance of a random sample of size 5 from a normal population with $\sigma^2 = 25$ will fall between 20 and 30.

78. The claim that the variance of a normal population is $\sigma^2 = 25$ is to be rejected if the variance of a random sample of size 16 exceeds 54.668 or is less than 12.102. What is the probability that this claim will be rejected even though $\sigma^2 = 25$?

79. The claim that the variance of a normal population is $\sigma^2 = 4$ is to be rejected if the variance of a random sample of size 9 exceeds 7.7535. What is the probability that this claim will be rejected even though $\sigma^2 = 4$?

80. A random sample of size $n = 25$ from a normal population has the mean $\bar{x} = 47$ and the standard deviation $s = 7$. If we base our decision on the statistic of Theorem 13, can we say that the given information supports the conjecture that the mean of the population is $\mu = 42$?

81. A random sample of size $n = 12$ from a normal population has the mean $\bar{x} = 27.8$ and the variance $s^2 = 3.24$. If we base our decision on the statistic of Theorem 13, can we say that the given information supports the claim that the mean of the population is $\mu = 28.5$?

82. If S_1 and S_2 are the standard deviations of independent random samples of sizes $n_1 = 61$ and $n_2 = 31$ from normal populations with $\sigma_1^2 = 12$ and $\sigma_2^2 = 18$, find $P(S_1^2/S_2^2 > 1.16)$.

83. If S_1^2 and S_2^2 are the variances of independent random samples of sizes $n_1 = 10$ and $n_2 = 15$ from normal populations with equal variances, find $P(S_1^2/S_2^2 < 4.03)$.

84. Use a computer program to verify the five entries in Table IV of "Statistical Tables" corresponding to 11 degrees of freedom.

85. Use a computer program to verify the eight entries in Table V of "Statistical Tables" corresponding to 21 degrees of freedom.

86. Use a computer program to verify the five values of $f_{0.05}$ in Table VI of "Statistical Tables" corresponding to 7 and 6 to 10 degrees of freedom.

87. Use a computer program to verify the six values of $f_{0.01}$ in Table VI of "Statistical Tables" corresponding to $\nu_1 = 15$ and $\nu_2 = 12, 13, \ldots, 17$.

SEC. 7

88. Find the probability that in a random sample of size $n = 4$ from the continuous uniform population of Exercise 46, the smallest value will be at least 0.20.

89. Find the probability that in a random sample of size $n = 3$ from the beta population of Exercise 77, the largest value will be less than 0.90.

90. Use the result of Exercise 56 to find the probability that the range of a random sample of size $n = 5$ from the given uniform population will be at least 0.75.

91. Use the result of part (c) of Exercise 58 to find the probability that in a random sample of size $n = 10$ at least 80 percent of the population will lie between the smallest and largest values.

92. Use the result of part (c) of Exercise 58 to set up an equation in n whose solution will give the sample size that is required to be able to assert with probability $1 - \alpha$ that the proportion of the population contained between the smallest and largest sample values is at least p. Show that for $p = 0.90$ and $\alpha = 0.05$ this equation can be written as

$$(0.90)^{n-1} = \frac{1}{2n + 18}$$

This kind of equation is difficult to solve, but it can be shown that an approximate solution for n is given by

$$\frac{1}{2} + \frac{1}{4} \cdot \frac{1+p}{1-p} \cdot \chi^2_{\alpha,4}$$

where $\chi^2_{\alpha,4}$ must be looked up in Table V of "Statistical Tables". Use this method to find an approximate solution of the equation for $p = 0.90$ and $\alpha = 0.05$.

SEC. 8

93. Cans of food, stacked in a warehouse, are sampled to determine the proportion of damaged cans. Explain why a sample that includes only the top can in each stack would not be a random sample.

94. An inspector chooses a sample of parts coming from an automated lathe by visually inspecting all parts, and then including 10 percent of the "good" parts in the sample with the use of a table of random digits.

(a) Why does this method not produce a random sample of the production of the lathe?

(b) Of what population can this be considered to be a random sample?

95. Sections of aluminum sheet metal of various lengths, used for construction of airplane fuselages, are lined up on a conveyor belt that moves at a constant speed. A sample is selected by taking whatever section is passing in front of a station at five-minute intervals. Explain why this sample may not be random; that is, it is not an accurate representation of the population of all aluminum sections.

96. A process error may cause the oxide thicknesses on the surface of a silicon wafer to be "wavy," with a constant difference between the wave heights. What precautions are necessary in taking a random sample of oxide thicknesses at various positions on the wafer to assure that the observations are independent?

References

Necessary and sufficient conditions for the strongest form of the central limit theorem for independent random variables, the *Lindeberg–Feller* conditions, are given in

Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. New York: John Wiley & Sons, Inc., 1968,

as well as in other advanced texts on probability theory.

Extensive tables of the normal, chi-square, F, and t distributions may be found in

Pearson, E. S., and Hartley, H. O., *Biometrika Tables for Statisticians*, Vol. I. New York: John Wiley & Sons, Inc., 1968.

A general formula for the variance of the sampling distribution of the second sample moment M_2 (which differs from S^2 only insofar as we divide by n instead of $n-1$) is derived in

Cramér, H., *Mathematical Methods of Statistics*. Princeton, N.J.: Princeton University Press, 1950,

and a proof of Theorem 17 is given in

Wilks, S. S., *Mathematical Statistics*. New York: John Wiley & Sons, Inc., 1962.

Proofs of the independence of $\overline{X}$ and S^2 for random samples from normal populations are given in many advanced texts on mathematical statistics. For instance, a proof based on moment-generating functions may be found in the above-mentioned book by S. S. Wilks, and a somewhat more elementary proof, illustrated for $n = 3$, may be found in

Keeping, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962.

The proof outlined in Exercise 48 is given in

Shuster, J., "A Simple Method of Teaching the Independence of $\overline{X}$ and S^2," *The American Statistician*, Vol. 27, No. 1, 1973.

Answers to Odd-Numbered Exercises

11 When we sample with replacement from a finite population, we satisfy the conditions for random sampling from an infinite population; that is, the random variables are independent and identically distributed.

17 $\mu = 13.0; \sigma^2 = 25.6$.

19 $s^2 = 4$.

29 21.9% and 5.53%.

47 $h(\tilde{x}) = \frac{(2m+1)!}{m!m!}\tilde{x}(1-\tilde{x})^m$ for $0 < x < 1$; $h(\tilde{x}) = 0$ elsewhere.

49 $g_1(y_1) = 12ny_1^2(1-y_1)(1-4y_1)^3$.

51 **(a)**

y_1	1	2	3	4
$g_1(y_1)$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

(b)

y_1	1	2	3	4	5
$g_1(y_1)$	$\frac{9}{25}$	$\frac{7}{25}$	$\frac{5}{25}$	$\frac{3}{25}$	$\frac{1}{25}$

53 $\frac{1}{(n+1)^2(n+2)}$.

55 $f(R) = \frac{n-1}{\theta}e^{-R/\theta}[1-e^{-R/\theta}]^{n-2}$ for $R > 0$; $f(R) = 0$ elsewhere.

57 $E(R) = \frac{n-1}{n+1}; \sigma^2 = \frac{2(n-1)}{(n+1)^2(n+2)}$.

61 **(a)** $\frac{1}{495}$; **(b)** $\frac{1}{77}$.

63 **(a)** It is divided by 2. **(b)** It is divided by 1.5. **(c)** It is multiplied by 3. **(d)** It is multiplied by 2.5.

65 **(a)** 0.96; **(b)** 0.9999994.

67 0.0250.

69 0.0207.

71 0.2302.

73 4.63.

75 **(a)** 0.3056; **(b)** 0.7698.

77 0.216.

79 0.5.

81 $t = -1.347$; the data support the claim.

83 0.99.

89 0.851.

91 0.6242.