Chapter 4

RANDOM VARIABLES AND EXPECTATION

4.1 RANDOM VARIABLES

When a random experiment is performed, we are often not interested in all of the details of the experimental result but only in the value of some numerical quantity determined by the result. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the values of the individual dice. That is, we may be interested in knowing that the sum is 7 and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). Also, a civil engineer may not be directly concerned with the daily risings and declines of the water level of a reservoir (which we can take as the experimental result) but may only care about the level at the end of a rainy season. These quantities of interest that are determined by the result of the experiment are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities of its possible values.

EXAMPLE 4.1a Letting X denote the random variable that is defined as the sum of two fair dice, then

$$P\{X = 2\} = P\{(1,1)\} = \tfrac{1}{36} \tag{4.1.1}$$

$$P\{X = 3\} = P\{(1,2),(2,1)\} = \tfrac{2}{36}$$

$$P\{X = 4\} = P\{(1,3),(2,2),(3,1)\} = \tfrac{3}{36}$$

$$P\{X = 5\} = P\{(1,4),(2,3),(3,2),(4,1)\} = \tfrac{4}{36}$$

$$P\{X = 6\} = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \tfrac{5}{36}$$

$$P\{X = 7\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \tfrac{6}{36}$$

$$P\{X = 8\} = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \tfrac{5}{36}$$

$$P\{X = 9\} = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = \tfrac{4}{36}$$

$$P\{X = 10\} = P\{(4, 6), (5, 5), (6, 4)\} = \tfrac{3}{36}$$

$$P\{X = 11\} = P\{(5, 6), (6, 5)\} = \tfrac{2}{36}$$

$$P\{X = 12\} = P\{(6, 6)\} = \tfrac{1}{36}$$

In other words, the random variable X can take on any integral value between 2 and 12 and the probability that it takes on each value is given by Equation 4.1.1. Since X must take on some value, we must have

$$1 = P(S) = P\left(\bigcup_{i=2}^{12}\{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

which is easily verified from Equation 4.1.1.

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

$$P\{Y = i\} = 1/6, \qquad i = 1, 2, 3, 4, 5, 6 \quad \blacksquare$$

EXAMPLE 4.1b   Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities .09, .21, .21, .49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the *indicator* random variable for the event A, since I will equal 1

or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I=1\} = .91$$
$$P\{I=0\} = .09 \quad \blacksquare$$

In the two foregoing examples, the random variables of interest took on a finite number of possible values. Random variables whose set of possible values can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$ are said to be *discrete*. For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The *cumulative distribution function*, or more simply the *distribution function*, F of the random variable X is defined for any real number x by

$$F(x) = P\{X \leq x\}$$

That is, $F(x)$ is the probability that the random variable X takes on a value that is less than or equal to x.

Notation: We will use the notation $X \sim F$ to signify that F is the distribution function of X.

All probability questions about X can be answered in terms of its distribution function F. For example, suppose we wanted to compute $P\{a < X \leq b\}$. This can be accomplished by first noting that the event $\{X \leq b\}$ can be expressed as the union of the two mutually exclusive events $\{X \leq a\}$ and $\{a < X \leq b\}$. Therefore, applying Axiom 3, we obtain that

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

or

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

EXAMPLE 4.1c   Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

SOLUTION   The desired probability is computed as follows:

$$P\{X > 1\} = 1 - P\{X \leq 1\}$$
$$= 1 - F(1)$$
$$= e^{-1}$$
$$= .368 \quad \blacksquare$$

4.2   TYPES OF RANDOM VARIABLES

As was previously mentioned, a random variable whose set of possible values is a sequence is said to be *discrete*. For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$p(x_i) > 0, \qquad i = 1, 2, \ldots$$
$$p(x) = 0, \qquad \text{all other values of } x$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

EXAMPLE 4.2a   Consider a random variable X that is equal to 1, 2, or 3. If we know that

$$p(1) = \tfrac{1}{2} \qquad \text{and} \qquad p(2) = \tfrac{1}{3}$$

then it follows (since $p(1) + p(2) + p(3) = 1$) that

$$p(3) = \tfrac{1}{6}$$

A graph of $p(x)$ is presented in Figure 4.1. $\blacksquare$

The cumulative distribution function F can be expressed in terms of $p(x)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function F is a step function. That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.
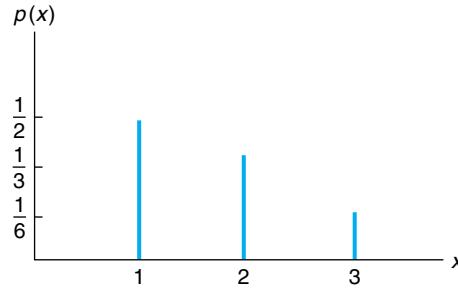
FIGURE 4.1   *Graph of p(x), Example 4.2a.*
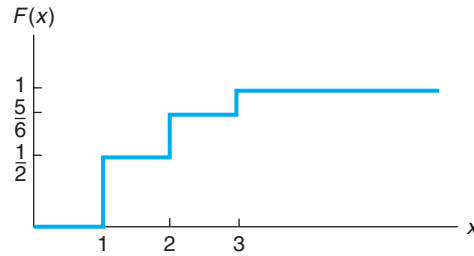
FIGURE 4.2   *Graph of F(x).*

For instance, suppose X has a probability mass function given (as in Example 4.2a) by

$$p(1) = \tfrac{1}{2}, \qquad p(2) = \tfrac{1}{3}, \qquad p(3) = \tfrac{1}{6}$$

Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \tfrac{1}{2} & 1 \le a < 2 \\ \tfrac{5}{6} & 2 \le a < 3 \\ 1 & 3 \le a \end{cases}$$

This is graphically presented in Figure 4.2.

Whereas the set of possible values of a discrete random variable is a sequence, we often must consider random variables whose set of possible values is an interval. Let X be such a random variable. We say that X is a *continuous* random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)\, dx \qquad (4.2.1)$$

The function $f(x)$ is called the *probability density function* of the random variable X.

In words, Equation 4.2.1 states that the probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, $f(x)$ must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\, dx$$

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain from Equation 4.2.1 that

$$P\{a \le X \le b\} = \int_{a}^{b} f(x)\, dx \qquad (4.2.2)$$

If we let $a = b$ in the above, then

$$P\{X = a\} = \int_{a}^{a} f(x)\, dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any *particular* value is zero. (See Figure 4.3.)

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides yields

$$\frac{d}{da} F(a) = f(a)$$

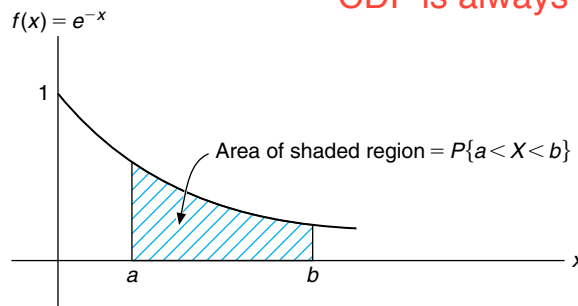CDF is always continuous and differentiable

$f(x) = e^{-x}$

1

Area of shaded region $= P\{a < X < b\}$

a b x

FIGURE 4.3   *The probability density function* $f(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & x < 0 \end{cases}$.

That is, the density is the derivative of the cumulative distribution function. A somewhat more intuitive interpretation of the density function may be obtained from Equation 4.2.2 as follows:

$$P\left\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\, dx \approx \varepsilon f(a)$$

when ε is small. In other words, the probability that X will be contained in an interval of length ε around the point a is approximately $\varepsilon f(a)$. From this, we see that $f(a)$ is a measure of how likely it is that the random variable will be near a.

EXAMPLE 4.2b   Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?
(b) Find $P\{X > 1\}$.

SOLUTION   (a) Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\, dx = 1$, implying that

$$C \int_0^2 (4x - 2x^2)\, dx = 1$$

or

$$C\left[2x^2 - \frac{2x^3}{3}\right]\Big|_{x=0}^{x=2} = 1$$

or

$$C = \tfrac{3}{8}$$

(b) Hence

$$P\{X > 1\} = \int_1^\infty f(x)\, dx = \tfrac{3}{8} \int_1^2 (4x - 2x^2)\, dx = \tfrac{1}{2} \quad \blacksquare$$

4.3   JOINTLY DISTRIBUTED RANDOM VARIABLES

For a given experiment, we are often interested not only in probability distribution functions of individual random variables but also in the relationships between two or more random variables. For instance, in an experiment into the possible causes of cancer, we

might be interested in the relationship between the average number of cigarettes smoked daily and the age at which an individual contracts cancer. Similarly, an engineer might be interested in the relationship between the shear strength and the diameter of a spot weld in a fabricated sheet steel specimen.

To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

A knowledge of the joint probability distribution function enables one, at least in theory, to compute the probability of any statement concerning the values of X and Y. For instance, the distribution function of X — call it F_X — can be obtained from the joint distribution function F of X and Y as follows:

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y < \infty\} \\ &= F(x, \infty) \end{aligned}$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(y) = F(\infty, y)$$

In the case where X and Y are both discrete random variables whose possible values are, respectively, $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$, we define the *joint probability mass function* of X and Y, $p(x_i, y_j)$, by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

The individual probability mass functions of X and Y are easily obtained from the joint probability mass function by the following reasoning. Since Y must take on some value y_j, it follows that the event $\{X = x_i\}$ can be written as the union, over all j, of the mutually exclusive events $\{X = x_i, Y = y_j\}$. That is,

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

and so, using Axiom 3 of the probability function, we see that

$$\begin{aligned} P\{X = x_i\} &= P\left(\bigcup_j \{X = x_i, Y = y_j\} \right) \qquad\qquad (4.3.1) \\ &= \sum_j P\{X = x_i, Y = y_j\} \\ &= \sum_j p(x_i, y_j) \end{aligned}$$

Similarly, we can obtain $P\{Y = y_j\}$ by summing $p(x_i, y_j)$ over all possible values of x_i, that is,

$$P\{Y = y_j\} = \sum_i P\{X = x_i, Y = y_j\} \qquad (4.3.2)$$
$$= \sum_i p(x_i, y_j)$$

Hence, specifying the joint probability mass function always determines the individual mass functions. However, it should be noted that the reverse is not true. Namely, knowledge of $P\{X = x_i\}$ and $P\{Y = y_j\}$ does not determine the value of $P\{X = x_i, Y = y_j\}$.

EXAMPLE 4.3a   Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability mass function of X and Y, $p(i, j) = P\{X = i, Y = j\}$, is given by

$$p(i, j) = \frac{\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}}{\binom{12}{3}}$$

where the preceding follows because of the $\binom{12}{3}$ equally likely outcomes, there are, by the basic principle of counting, $\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}$ possible choices that contain exactly i new, j used, and $3 - i - j$ defective batteries. Consequently,

$$p(0, 0) = \binom{5}{3} \Big/ \binom{12}{3} = 10/220$$

$$p(0, 1) = \binom{4}{1}\binom{5}{2} \Big/ \binom{12}{3} = 40/220$$

$$p(0, 2) = \binom{4}{2}\binom{5}{1} \Big/ \binom{12}{3} = 30/220$$

$$p(0, 3) = \binom{4}{3} \Big/ \binom{12}{3} = 4/220$$

$$p(1, 0) = \binom{3}{1}\binom{5}{2} \Big/ \binom{12}{3} = 30/220$$

$$p(1, 1) = \binom{3}{1}\binom{4}{1}\binom{5}{1} \Big/ \binom{12}{3} = 60/220$$

$$p(1, 2) = \binom{3}{1}\binom{4}{2} \Big/ \binom{12}{3} = 18/220$$

$$p(2, 0) = \binom{3}{2}\binom{5}{1} \Big/ \binom{12}{3} = 15/220$$

TABLE 4.1 $P\{X = i, Y = j\}$

i \ j	0	1	2	3	Row Sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums = $P\{Y=j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

$$p(2, 1) = \binom{3}{2}\binom{4}{1}\bigg/\binom{12}{3} = 12/220$$

$$p(3, 0) = \binom{3}{3}\bigg/\binom{12}{3} = 1/220$$

These probabilities can most easily be expressed in tabular form as shown in Table 4.1.

The reader should note that the probability mass function of X is obtained by computing the row sums, in accordance with the Equation 4.3.1, whereas the probability mass function of Y is obtained by computing the column sums, in accordance with Equation 4.3.2. Because the individual probability mass functions of X and Y thus appear in the margin of such a table, they are often referred to as being the marginal probability mass functions of X and Y, respectively. It should be noted that to check the correctness of such a table we could sum the marginal row (or the marginal column) and verify that its sum is 1. (Why must the sum of the entries in the marginal row (or column) equal 1?)   ■

EXAMPLE 4.3b   Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1, 35 percent have 2, and 30 percent have 3 children; suppose further that each child is equally likely (and independently) to be a boy or a girl. If a family is chosen at random from this community, then B, the number of boys, and G, the number of girls, in this family will have the joint probability mass function shown in Table 4.2.

These probabilities are obtained as follows:

$$P\{B = 0, G = 0\} = P\{\text{no children}\}$$
$$= .15$$
$$P\{B = 0, G = 1\} = P\{\text{1 girl and total of 1 child}\}$$
$$= P\{\text{1 child}\}P\{\text{1 girl}|\text{1 child}\}$$
$$= (.20)\left(\tfrac{1}{2}\right) = .1$$

TABLE 4.2 $P\{B = i, G = j\}$

i $\diagdown$ j	0	1	2	3	Row Sum = $P\{B = i\}$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
Column Sum = $P\{G = j\}$	.3750	.3875	.2000	.0375	

$$P\{B = 0, G = 2\} = P\{2 \text{ girls and total of 2 children}\}$$
$$= P\{2 \text{ children}\}P\{2 \text{ girls}|2 \text{ children}\}$$
$$= (.35)\left(\tfrac{1}{2}\right)^2 = .0875$$
$$P\{B = 0, G = 3\} = P\{3 \text{ girls and total of 3 children}\}$$
$$= P\{3 \text{ children}\}P\{3 \text{ girls}|3 \text{ children}\}$$
$$= (.30)\left(\tfrac{1}{2}\right)^3 = .0375$$

We leave it to the reader to verify the remainder of Table 4.2, which tells us, among other things, that the family chosen will have at least 1 girl with probability .625. ∎

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane)

$$P\{(X, Y) \in C\} = \iint\limits_{(x,y)\in C} f(x, y) \, dx \, dy \tag{4.3.3}$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from Equation 4.3.3 that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y) \, dx \, dy \tag{4.3.4}$$

Because

$$F(a, b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\}$$
$$= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y) \, dx \, dy$$

it follows, upon differentiation, that

$$f(a, b) = \frac{\partial^2}{\partial a \, \partial b} F(a, b)$$

wherever the partial derivatives are defined. Another interpretation of the joint density function is obtained from Equation 4.3.4 as follows:

$$P\{a < X < a + da, b < Y < b + db\} = \int_b^{d+db} \int_a^{a+da} f(x, y) \, dx \, dy$$
$$\approx f(a, b) \, da \, db$$

when da and db are small and $f(x, y)$ is continuous at a, b. Hence $f(a, b)$ is a measure of how likely it is that the random vector (X, Y) will be near (a, b).

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\} \qquad (4.3.5)$$

$$= \int_A \int_{-\infty}^{\infty} f(x, y) \, dy \, dx$$

$$= \int_A f_X(x) \, dx$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy$$

is thus the probability density function of X. Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx \qquad (4.3.6)$$

EXAMPLE 4.3c   The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute **(a)** $P\{X > 1, Y < 1\}$; **(b)** $P\{X < Y\}$; and **(c)** $P\{X < a\}$.

SOLUTION

(a)
$$P\{X > 1, Y < 1\} = \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\, dx\, dy$$

$$= \int_0^1 2e^{-2y}(-e^{-x}|_1^\infty)\, dy$$

$$= e^{-1} \int_0^1 2e^{-2y}\, dy$$

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

(b)
$$P\{X < Y\} = \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\, dx\, dy$$

$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\, dx\, dy$$

$$= \int_0^\infty 2e^{-2y}(1 - e^{-y})\, dy$$

$$= \int_0^\infty 2e^{-2y}\, dy - \int_0^\infty 2e^{-3y}\, dy$$

$$= 1 - \tfrac{2}{3}$$

$$= \tfrac{1}{3}$$

(c)
$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-2y}e^{-x}\, dy\, dx$$

$$= \int_0^a e^{-x}\, dx$$

$$= 1 - e^{-a} \quad \blacksquare$$

4.3.1 INDEPENDENT RANDOM VARIABLES

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \tag{4.3.7}$$

In other words, X and Y are independent if, for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.

It can be shown by using the three axioms of probability that Equation 4.3.7 will follow if and only if for all a, b

$$P\{X \le a, Y \le b\} = P\{X \le a\}P\{Y \le b\}$$

Hence, in terms of the joint distribution function F of X and Y, we have that X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \qquad \text{for all } a, b$$

When X and Y are discrete random variables, the condition of independence Equation 4.3.7 is equivalent to

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{for all } x, y \tag{4.3.8}$$

where p_X and p_Y are the probability mass functions of X and Y. The equivalence follows because, if Equation 4.3.7 is satisfied, then we obtain Equation 4.3.8 by letting A and B be, respectively, the one-point sets $A = \{x\}, B = \{y\}$. Furthermore, if Equation 4.3.8 is valid, then for any sets A, B

$$
\begin{aligned}
P\{X \in A, Y \in B\} &= \sum_{y \in B}\sum_{x \in A} p(x, y) \\
&= \sum_{y \in B}\sum_{x \in A} p_X(x)p_Y(y) \\
&= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) \\
&= P\{Y \in B\}P\{X \in A\}
\end{aligned}
$$

and thus Equation 4.3.7 is established.

In the jointly continuous case, the condition of independence is equivalent to

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x, y$$

Loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent.

EXAMPLE 4.3d   Suppose that X and Y are independent random variables having the common density function

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

Find the density function of the random variable X/Y.

SOLUTION   We start by determining the distribution function of X/Y. For $a > 0$

$$F_{X/Y}(a) = P\{X/Y \le a\}$$

$$= \iint_{x/y \le a} f(x, y) \, dx \, dy$$

$$= \iint_{x/y \le a} e^{-x} e^{-y} \, dx \, dy$$

$$= \int_0^\infty \int_0^{ay} e^{-x} e^{-y} \, dx \, dy$$

$$= \int_0^\infty (1 - e^{-ay}) e^{-y} \, dy$$

$$= \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right] \Big|_0^\infty$$

$$= 1 - \frac{1}{a+1}$$

Differentiation yields that the density function of X/Y is given by

$$f_{X/Y}(a) = 1/(a+1)^2, \qquad 0 < a < \infty \quad \blacksquare$$

We can also define joint probability distributions for n random variables in exactly the same manner as we did for $n = 2$. For instance, the joint cumulative probability distribution function $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n\}$$

If these random variables are discrete, we define their joint probability mass function $p(x_1, x_2, \ldots, x_n)$ by

$$p(x_1, x_2, \ldots, x_n) = P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$$

Further, the n random variables are said to be jointly continuous if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the joint probability density function, such that for any set C in n-space

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \int \int_{(x_1, \ldots, x_n) \in C} \cdots \int f(x_1, \ldots, x_n) \, dx_1 \, dx_2 \cdots dx_n$$

In particular, for any n sets of real numbers $A_1, A_2, \ldots, A_n$

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\}$$
$$= \int_{A_n} \int_{A_{n-1}} \cdots \int_{A_1} f(x_1, \ldots, x_n) \, dx_1 \, dx_2 \ldots dx_n$$

The concept of independence may, of course, also be defined for more than two random variables. In general, the n random variables $X_1, X_2, \ldots, X_n$ are said to be independent if, for all sets of real numbers $A_1, A_2, \ldots, A_n$,

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

As before, it can be shown that this condition is equivalent to

$$P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$
$$= \prod_{i=1}^{n} P\{X_1 \leq a_i\} \qquad \text{for all } a_1, a_2, \ldots, a_n$$

Finally, we say that an infinite collection of random variables is independent if every finite subcollection of them is independent.

EXAMPLE 4.3e Suppose that the successive daily changes of the price of a given stock are assumed to be independent and identically distributed random variables with probability mass function given by

$$P\{\text{daily change is } i\} = \begin{cases} -3 & \text{with probability } .05 \\ -2 & \text{with probability } .10 \\ -1 & \text{with probability } .20 \\ 0 & \text{with probability } .30 \\ 1 & \text{with probability } .20 \\ 2 & \text{with probability } .10 \\ 3 & \text{with probability } .05 \end{cases}$$

Then the probability that the stock's price will increase successively by 1, 2, and 0 points in the next three days is

$$P\{X_1 = 1, X_2 = 2, X_3 = 0\} = (.20)(.10)(.30) = .006$$

where we have let X_i denote the change on the ith day. ■

*4.3.2 CONDITIONAL DISTRIBUTIONS

The relationship between two random variables can often be clarified by consideration of the conditional distribution of one given the value of the other.

Recall that for any two events E and F, the conditional probability of E given F is defined, provided that $P(F) > 0$, by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence, if X and Y are discrete random variables, it is natural to define the conditional probability mass function of X given that $Y = y$, by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x,y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $p_Y(y) > 0$.

EXAMPLE 4.3f   If we know, in Example 4.3b, that the family chosen has one girl, compute the conditional probability mass function of the number of boys in the family.

SOLUTION   We first note from Table 4.2 that

$$P\{G = 1\} = .3875$$

Hence,

$$P\{B = 0|G = 1\} = \frac{P\{B = 0, G = 1\}}{P\{G = 1\}} = \frac{.10}{.3875} = 8/31$$

$$P\{B = 1|G = 1\} = \frac{P\{B = 1, G = 1\}}{P\{G = 1\}} = \frac{.175}{.3875} = 14/31$$

$$P\{B = 2|G = 1\} = \frac{P\{B = 2, G = 1\}}{P\{G = 1\}} = \frac{.1125}{.3875} = 9/31$$

$$P\{B = 3|G = 1\} = \frac{P\{B = 3, G = 1\}}{P\{G = 1\}} = 0$$

Thus, for instance, given 1 girl, there are 23 chances out of 31 that there will also be at least 1 boy. ■

---

* Optional section.

EXAMPLE 4.3g   Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by

$$p(0, 0) = .4, \quad p(0, 1) = .2, \quad p(1, 0) = .1, \quad p(1, 1) = .3$$

Calculate the conditional probability mass function of X given that $Y = 1$.

SOLUTION   We first note that

$$P\{Y = 1\} = \sum_x p(x, 1) = p(0, 1) + p(1, 1) = .5$$

Hence,

$$P\{X = 0 | Y = 1\} = \frac{p(0, 1)}{P\{Y = 1\}} = 2/5$$

$$P\{X = 1 | Y = 1\} = \frac{p(1, 1)}{P\{Y = 1\}} = 3/5 \quad \blacksquare$$

If X and Y have a joint probability density function $f(x, y)$, then the conditional probability density function of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}$$

To motivate this definition, multiply the left-hand side by dx and the right-hand side by $(dx\, dy)/dy$ to obtain

$$
\begin{aligned}
f_{X|Y}(x | y)\, dx &= \frac{f(x, y)\, dx\, dy}{f_Y(y)\, dy} \\
&\approx \frac{P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}}{P\{y \leq Y \leq y + dy\}} \\
&= P\{x \leq X \leq x + dy | y \leq Y \leq y + dy\}
\end{aligned}
$$

In other words, for small values of dx and dy, $f_{X|Y}(x | y)\, dx$ represents the conditional probability that X is between x and $x + dx$, given that Y is between y and $y + dy$.

The use of conditional densities allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable. That is, if X and Y are jointly continuous, then, for any set A,

$$P\{X \in A | Y = y\} = \int_A f_{X|Y}(x | y)\, dx$$

EXAMPLE 4.3h The joint density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{12}{5}x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional density of X, given that $Y = y$, where $0 < y < 1$.

SOLUTION For $0 < x < 1, 0 < y < 1$, we have

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

$$= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, dx}$$

$$= \frac{x(2 - x - y)}{\int_0^1 x(2 - x - y)\, dx}$$

$$= \frac{x(2 - x - y)}{\frac{2}{3} - y/2}$$

$$= \frac{6x(2 - x - y)}{4 - 3y} \quad \blacksquare$$

4.4 EXPECTATION

One of the most important concepts in probability theory is that of the expectation of a random variable. If X is a discrete random variable taking on the possible values $x_1, x_2, \ldots,$ then the *expectation* or *expected value* of X, denoted by $E[X]$, is defined by

$$E[X] = \sum_i x_i P\{X = x_i\}$$

In words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it. For instance, if the probability mass function of X is given by

$$p(0) = \tfrac{1}{2} = p(1)$$

then

$$E[X] = 0\left(\tfrac{1}{2}\right) + 1\left(\tfrac{1}{2}\right) = \tfrac{1}{2}$$

is just the ordinary average of the two possible values 0 and 1 that X can assume. On the other hand, if

$$p(0) = \tfrac{1}{3}, \quad p(1) = \tfrac{2}{3}$$

then

$$E[X] = 0 \left(\tfrac{1}{3}\right) + 1 \left(\tfrac{2}{3}\right) = \tfrac{2}{3}$$

is a weighted average of the two possible values 0 and 1 where the value 1 is given twice as much weight as the value 0 since $p(1) = 2p(0)$.

Another motivation of the definition of expectation is provided by the frequency interpretation of probabilities. This interpretation assumes that if an infinite sequence of independent replications of an experiment is performed, then for any event E, the proportion of time that E occurs will be $P(E)$. Now, consider a random variable X that must take on one of the values $x_1, x_2, \ldots, x_n$ with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$; and think of X as representing our winnings in a single game of chance. That is, with probability $p(x_i)$ we shall win x_i units $i = 1, 2, \ldots, n$. Now by the frequency interpretation, it follows that if we continually play this game, then the proportion of time that we win x_i will be $p(x_i)$. Since this is true for all i, $i = 1, 2, \ldots, n$, it follows that our average winnings per game will be

$$\sum_{i=1}^{n} x_i p(x_i) = E[X]$$

To see this argument more clearly, suppose that we play N games where N is very large. Then in approximately $Np(x_i)$ of these games, we shall win x_i, and thus our total winnings in the N games will be

$$\sum_{i=1}^{n} x_i N\, p(x_i)$$

implying that our average winnings per game are

$$\sum_{i=1}^{n} \frac{x_i Np(x_i)}{N} = \sum_{i=1}^{n} x_i p(x_i) = E[X]$$

EXAMPLE 4.4a   Find $E[X]$ where X is the outcome when we roll a fair die.

SOLUTION   Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \tfrac{1}{6}$, we obtain that

$$E[X] = 1 \left(\tfrac{1}{6}\right) + 2 \left(\tfrac{1}{6}\right) + 3 \left(\tfrac{1}{6}\right) + 4 \left(\tfrac{1}{6}\right) + 5 \left(\tfrac{1}{6}\right) + 6 \left(\tfrac{1}{6}\right) = \tfrac{7}{2}$$

The reader should note that, for this example, the expected value of X is not a value that X could possibly assume. (That is, rolling a die cannot possibly lead to an outcome of 7/2.) Thus, even though we call $E[X]$ the *expectation* of X, it should not be interpreted as the value that we *expect* X to have but rather as the average value of X in a large number of

repetitions of the experiment. That is, if we continually roll a fair die, then after a large number of rolls the average of all the outcomes will be approximately 7/2. (The interested reader should try this as an experiment.) ■

EXAMPLE 4.4b   If I is an indicator random variable for the event A, that is, if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

then

$$E[I] = 1P(A) + 0P(A^c) = P(A)$$

Hence, the expectation of the indicator random variable for the event A is just the probability that A occurs. ■

EXAMPLE 4.4c (Entropy)   For a given random variable X, how much information is conveyed in the message that $X = x$? Let us begin our attempts at quantifying this statement by agreeing that the amount of information in the message that $X = x$ should depend on how likely it was that X would equal x. In addition, it seems reasonable that the more unlikely it was that X would equal x, the more informative would be the message. For instance, if X represents the sum of two fair dice, then there seems to be more information in the message that X equals 12 than there would be in the message that X equals 7, since the former event has probability $\frac{1}{36}$ and the latter $\frac{1}{6}$.

Let us denote by $I(p)$ the amount of information contained in the message that an event, whose probability is p, has occurred. Clearly $I(p)$ should be a nonnegative, decreasing function of p. To determine its form, let X and Y be independent random variables, and suppose that $P\{X = x\} = p$ and $P\{Y = y\} = q$. How much information is contained in the message that X equals x and Y equals y? To answer this, note first that the amount of information in the statement that X equals x is $I(p)$. Also, since knowledge of the fact that X is equal to x does not affect the probability that Y will equal y (since X and Y are independent), it seems reasonable that the additional amount of information contained in the statement that $Y = y$ should equal $I(q)$. Thus, it seems that the amount of information in the message that X equals x and Y equals y is $I(p) + I(q)$. On the other hand, however, we have that

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\} = pq$$

which implies that the amount of information in the message that X equals x and Y equals y is $I(pq)$. Therefore, it seems that the function I should satisfy the identity

$$I(pq) = I(p) + I(q)$$

However, if we define the function G by

$$G(p) = I(2^{-p})$$

then we see from the above that

$$\begin{aligned}
G(p+q) &= I(2^{-(p+q)}) \\
&= I(2^{-p}2^{-q}) \\
&= I(2^{-p}) + I(2^{-q}) \\
&= G(p) + G(q)
\end{aligned}$$

However, it can be shown that the only (monotone) functions G that satisfy the foregoing functional relationship are those of the form

$$G(p) = cp$$

for some constant c. Therefore, we must have that

$$I(2^{-p}) = cp$$

or, letting $q = 2^{-p}$

$$I(q) = -c \log_2(q)$$

for some positive constant c. It is traditional to let $c = 1$ and to say that the information is measured in units of *bits* (short for binary digits).

Consider now a random variable X, which must take on one of the values $x_1, \ldots, x_n$ with respective probabilities $p_1, \ldots, p_n$. As $-\log_2(p_i)$ represents the information conveyed by the message that X is equal to x_i, it follows that the expected amount of information that will be conveyed when the value of X is transmitted is given by

$$H(X) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

The quantity $H(X)$ is known in information theory as the *entropy* of the random variable X. ■

We can also define the expectation of a continuous random variable. Suppose that X is a continuous random variable with probability density function f. Since, for dx small

$$f(x)\, dx \approx P\{x < X < x + dx\}$$

it follows that a weighted average of all possible values of X, with the weight given to x equal to the probability that X is near x, is just the integral over all x of $xf(x)\, dx$. Hence,

−1   0   1   2
$p(-1) = .10$, $p(0) = .25$, $p(1) = .30$, $p(2) = .35$
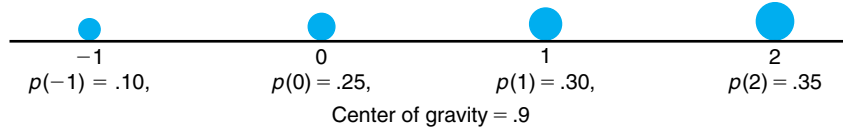Center of gravity = .9

FIGURE 4.4

it is natural to define the expected value of X by

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

EXAMPLE 4.4d Suppose that you are expecting a message at some time past 5 P.M. From experience you know that X, the number of hours after 5 P.M. until the message arrives, is a random variable with the following probability density function:

$$f(x) = \begin{cases} \dfrac{1}{1.5} & \text{if } 0 < x < 1.5 \\ 0 & \text{otherwise} \end{cases}$$

The expected amount of time past 5 P.M. until the message arrives is given by

$$E[X] = \int_0^{1.5} \frac{x}{1.5}\, dx = .75$$

Hence, on average, you would have to wait three-fourths of an hour. ■

REMARKS

(a) The concept of expectation is analogous to the physical concept of the center of gravity of a distribution of mass. Consider a discrete random variable X having probability mass function $p(x_i), i \geq 1$. If we now imagine a weightless rod in which weights with mass $p(x_i), i \geq 1$ are located at the points $x_i, i \geq 1$ (see Figure 4.4), then the point at which the rod would be in balance is known as the center of gravity. For those readers acquainted with elementary statics, it is now a simple matter to show that this point is at $E[X]$.[*]
(b) $E[X]$ has the same units of measurement as does X.

4.5 PROPERTIES OF THE EXPECTED VALUE

Suppose now that we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case). Suppose also that we are interested in calculating, not the expected

---

[*] To prove this, we must show that the sum of the torques tending to turn the point around $E[X]$ is equal to 0. That is, we must show that $0 = \sum_i (x_i - E[X]) p(x_i)$, which is immediate.

value of X, but the expected value of some function of X, say $g(X)$. How do we go about doing this? One way is as follows. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.

EXAMPLE 4.5a   Suppose X has the following probability mass function

$$p(0) = .2, \quad p(1) = .5, \quad p(2) = .3$$

Calculate $E[X^2]$.

SOLUTION   Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities

$$p_Y(0) = P\{Y = 0^2\} = .2$$
$$p_Y(1) = P\{Y = 1^2\} = .5$$
$$p_Y(4) = P\{Y = 2^2\} = .3$$

Hence,

$$E[X^2] = E[Y] = 0(.2) + 1(.5) + 4(.3) = 1.7 \quad \blacksquare$$

EXAMPLE 4.5b   The time, in hours, it takes to locate and repair an electrical break-down in a certain factory is a random variable — call it X — whose density function is given by

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

If the cost involved in a breakdown of duration x is x^3, what is the expected cost of such a breakdown?

SOLUTION   Letting $Y = X^3$ denote the cost, we first calculate its distribution function as follows. For $0 \leq a \leq 1$,

$$F_Y(a) = P\{Y \leq a\}$$
$$= P\{X^3 \leq a\}$$
$$= P\{X \leq a^{1/3}\}$$
$$= \int_0^{a^{1/3}} dx$$
$$= a^{1/3}$$

By differentiating $F_Y(a)$, we obtain the density of Y,

$$f_Y(a) = \frac{1}{3}a^{-2/3}, \quad 0 \le a < 1$$

Hence,

$$E[X^3] = E[Y] = \int_{-\infty}^{\infty} a f_Y(a)\, da$$

$$= \int_0^1 a\frac{1}{3}a^{-2/3}\, da$$

$$= \frac{1}{3}\int_0^1 a^{1/3}\, da$$

$$= \frac{1}{3}\frac{3}{4}a^{4/3}\big|_0^1$$

$$= \frac{1}{4} \quad \blacksquare$$

While the foregoing procedure will, in theory, always enable us to compute the expectation of any function of X from a knowledge of the distribution of X, there is an easier way of doing this. Suppose, for instance, that we wanted to compute the expected value of $g(X)$. Since $g(X)$ takes on the value $g(x)$ when $X = x$, it seems intuitive that $E[g(X)]$ should be a weighted average of the possible values $g(x)$ with, for a given x, the weight given to $g(x)$ being equal to the probability (or probability density in the continuous case) that X will equal x. Indeed, the foregoing can be shown to be true and we thus have the following proposition.

PROPOSITION 4.5.1 EXPECTATION OF A FUNCTION OF A RANDOM VARIABLE

(a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

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

(b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

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

EXAMPLE 4.5c Applying Proposition 4.5.1 to Example 4.5a yields

$$E[X^2] = 0^2(0.2) + (1^2)(0.5) + (2^2)(0.3) = 1.7$$

which, of course, checks with the result derived in Example 4.5a. ■

EXAMPLE 4.5d Applying the proposition to Example 4.5b yields

$$E[X^3] = \int_0^1 x^3 dx \quad (\text{since } f(x) = 1, 0 < x < 1)$$

$$= \frac{1}{4} \quad ■$$

An immediate corollary of Proposition 4.5.1 is the following.

Corollary 4.5.2

If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

Proof

In the discrete case,

$$E[aX + b] = \sum_x (ax + b)p(x)$$

$$= a\sum_x x\,p(x) + b\sum_x p(x)$$

$$= aE[X] + b$$

In the continuous case,

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)\,dx$$

$$= a\int_{-\infty}^{\infty} xf(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx$$

$$= aE[X] + b \quad ■$$

If we take $a = 0$ in Corollary 4.5.2, we see that

$$E[b] = b$$

That is, the expected value of a constant is just its value. (Is this intuitive?) Also, if we take $b = 0$, then we obtain

$$E[aX] = aE[X]$$

or, in words, the expected value of a constant multiplied by a random variable is just the constant times the expected value of the random variable. The expected value of a random variable X, $E[X]$, is also referred to as the *mean* or the *first moment* of X. The quantity $E[X^n]$, $n \geq 1$, is called the nth moment of X. By Proposition 4.5.1, we note that

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

4.5.1 EXPECTED VALUE OF SUMS OF RANDOM VARIABLES

The two-dimensional version of Proposition 4.5.1 states that if X and Y are random variables and g is a function of two variables, then

$$E[g(X, Y)] = \sum_y \sum_x g(x, y) p(x, y) \qquad \text{in the discrete case}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy \quad \text{in the continuous case}$$

For example, if $g(X, Y) = X + Y$, then, in the continuous case,

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

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y)\, dx\, dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y)\, dx\, dy$$

$$= E[X] + E[Y]$$

where the final equality followed by applying the identity

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

first to the function $g(x, y) = x$ and then to the function $g(x, y) = y$.

   A similar result can be shown in the discrete case and indeed, for any random variables X and Y,

$$E[X + Y] = E[X] + E[Y] \tag{4.5.1}$$

By repeatedly applying Equation 4.5.1 we can show that the expected value of the sum of any number of random variables equals the sum of their individual expectations. For instance,

$$
\begin{aligned}
E[X + Y + Z] &= E[(X + Y) + Z] \\
&= E[X + Y] + E[Z] \qquad \text{by Equation 4.5.1} \\
&= E[X] + E[Y] + E[Z] \qquad \text{again by Equation 4.5.1}
\end{aligned}
$$

And in general, for any n,

$$
E[X_1 + X_2 \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] \tag{4.5.2}
$$

Equation 4.5.2 is an extremely useful formula whose utility will now be illustrated by a series of examples.

EXAMPLE 4.5e   Find the expected value of the sum obtained when two fair dice are rolled.

SOLUTION   If X is the sum, then $E[X]$ can be obtained from the formula

$$
E[X] = \sum_{i=2}^{12} i\, P(X = i)
$$

However, it is simpler to name the dice, and let X_i be the value on dice $i, i = 1, 2$. As, $X = X_1 + X_2$, this yields that

$$
E[X] = E[X_1] + E[X_2]
$$

Thus, from Example 4.4a, we see that $E[X] = 7$.   ■

EXAMPLE 4.5f   A construction firm has recently sent in bids for 3 jobs worth (in profits) 10, 20, and 40 (thousand) dollars. If its probabilities of winning the jobs are respectively .2, .8, and .3, what is the firm's expected total profit?

SOLUTION   Letting $X_i, i = 1, 2, 3$ denote the firm's profit from job i, then

$$
\text{total profit} = X_1 + X_2 + X_3
$$

and so

$$
E[\text{total profit}] = E[X_1] + E[X_2] + E[X_3]
$$

Now

$$
\begin{aligned}
E[X_1] &= 10(.2) + 0(.8) = 2 \\
E[X_2] &= 20(.8) + 0(.2) = 16 \\
E[X_3] &= 40(.3) + 0(.7) = 12
\end{aligned}
$$

and thus the firm's expected total profit is 30 thousand dollars.   ■

EXAMPLE 4.5g A secretary has typed N letters along with their respective envelopes. The envelopes get mixed up when they fall on the floor. If the letters are placed in the mixed-up envelopes in a completely random manner (that is, each letter is equally likely to end up in any of the envelopes), what is the expected number of letters that are placed in the correct envelopes?

SOLUTION Letting X denote the number of letters that are placed in the correct envelope, we can most easily compute $E[X]$ by noting that

$$X = X_1 + X_2 + \cdots + X_N$$

where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th letter is placed in its proper envelope} \\ 0 & \text{otherwise} \end{cases}$$

Now, since the ith letter is equally likely to be put in any of the N envelopes, it follows that

$$P\{X_i = 1\} = P\{i\text{th letter is in its proper envelope}\} = 1/N$$

and so

$$E[X_i] = 1P\{X_i = 1\} + 0P\{X_i = 0\} = 1/N$$

Hence, from Equation 4.5.2 we obtain that

$$E[X] = E[X_1] + \cdots + E[X_N] = \left(\frac{1}{N}\right) N = 1$$

Hence, no matter how many letters there are, on the average, exactly one of the letters will be in its own envelope. ∎

EXAMPLE 4.5h Suppose there are 20 different types of coupons and suppose that each time one obtains a coupon it is equally likely to be any one of the types. Compute the expected number of different types that are contained in a set for 10 coupons.

SOLUTION Let X denote the number of different types in the set of 10 coupons. We compute $E[X]$ by using the representation

$$X = X_1 + \cdots + X_{20}$$

where

$$X_i = \begin{cases} 1 & \text{if at least one type } i \text{ coupon is contained in the set of 10} \\ 0 & \text{otherwise} \end{cases}$$

Now

$$E[X_i] = P\{X_i = 1\}$$
$$= P\{\text{at least one type } i \text{ coupon is in the set of 10}\}$$
$$= 1 - P\{\text{no type } i \text{ coupons are contained in the set of 10}\}$$
$$= 1 - \left(\tfrac{19}{20}\right)^{10}$$

when the last equality follows since each of the 10 coupons will (independently) not be a type i with probability $\frac{19}{20}$. Hence,

$$E[X] = E[X_1] + \cdots + E[X_{20}] = 20\left[1 - \left(\tfrac{19}{20}\right)^{10}\right] = 8.025 \quad \blacksquare$$

An important property of the mean arises when one must predict the value of a random variable. That is, suppose that the value of a random variable X is to be predicted. If we predict that X will equal c, then the square of the "error" involved will be $(X - c)^2$. We will now show that the average squared error is minimized when we predict that X will equal its mean μ. To see this, note that for any constant c

$$E[(X - c)^2] = E[(X - \mu + \mu - c)^2]$$
$$= E[(X - \mu)^2 + 2(\mu - c)(X - \mu) + (\mu - c)^2]$$
$$= E[(X - \mu)^2] + 2(\mu - c)E[X - \mu] + (\mu - c)^2$$
$$= E[(X - \mu)^2] + (\mu - c)^2 \quad \text{since} \quad E[X - \mu] = E[X] - \mu = 0$$
$$\geq E[(X - \mu)^2]$$

Hence, the best predictor of a random variable, in terms of minimizing the expected square of its error, is just its mean.

4.6 VARIANCE

Given a random variable X along with its probability distribution function, it would be extremely useful if we were able to summarize the essential properties of the mass function by certain suitably defined measures. One such measure would be $E[X]$, the expected value of X. However, while $E[X]$ yields the weighted average of the possible values of X, it does not tell us anything about the variation, or spread, of these values. For instance, while the following random variables W, Y, and Z having probability mass functions determined by

$$W = 0 \quad \text{with probability 1}$$

$$Y = \begin{cases} -1 & \text{with probability } \tfrac{1}{2} \\ 1 & \text{with probability } \tfrac{1}{2} \end{cases}$$

$$Z = \begin{cases} -100 & \text{with probability } \frac{1}{2} \\ 100 & \text{with probability } \frac{1}{2} \end{cases}$$

all have the same expectation — namely, 0 — there is much greater spread in the possible values of Y than in those of W (which is a constant) and in the possible values of Z than in those of Y.

Because we expect X to take on values around its mean $E[X]$, it would appear that a reasonable way of measuring the possible variation of X would be to look at how far apart X would be from its mean on the average. One possible way to measure this would be to consider the quantity $E[|X - \mu|]$, where $\mu = E[X]$, and $|X - \mu|$ represents the absolute value of $X - \mu$. However, it turns out to be mathematically inconvenient to deal with this quantity and so a more tractable quantity is usually considered — namely, the expectation of the square of the difference between X and its mean. We thus have the following definition.

Definition

If X is a random variable with mean μ, then the *variance* of X, denoted by Var(X), is defined by

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

An alternative formula for Var(X) can be derived as follows:

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

That is,

$$\text{Var}(X) = E[X^2] - (E[X])^2 \qquad (4.6.1)$$

or, in words, the variance of X is equal to the expected value of the square of X minus the square of the expected value of X. This is, in practice, often the easiest way to compute Var(X).

EXAMPLE 4.6a   Compute $\text{Var}(X)$ when X represents the outcome when we roll a fair die.

SOLUTION   Since $P\{X = i\} = \frac{1}{6}, i = 1, 2, 3, 4, 5, 6$, we obtain

$$E[X^2] = \sum_{i-1}^{6} i^2 P\{X = i\}$$
$$= 1^2 \left(\tfrac{1}{6}\right) + 2^2 \left(\tfrac{1}{6}\right) + 3^2 \left(\tfrac{1}{6}\right) + 4^2 \left(\tfrac{1}{6}\right) + 5^2 \left(\tfrac{1}{6}\right) + 6^2 \left(\tfrac{1}{6}\right)$$
$$= \tfrac{91}{6}$$

Hence, since it was shown in Example 4.4a that $E[X] = \frac{7}{2}$, we obtain from Equation 4.6.1 that

$$\text{Var}(X) = E[X^2] - (E[X])^2$$
$$= \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2 = \tfrac{35}{12} \quad \blacksquare$$

EXAMPLE 4.6b   Variance of an Indicator Random Variable. If, for some event A,

$$I = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{if event } A \text{ does not occur} \end{cases}$$

then

$$\text{Var}(I) = E[I^2] - (E[I])^2$$
$$= E[I] - (E[I])^2 \quad \text{since } I^2 = I \text{ (as } 1^2 = 1 \text{ and } 0^2 = 0)$$
$$= E[I](1 - E[I])$$
$$= P(A)[1 - P(A)] \quad \text{since } E[I] = P(A) \text{ from Example 4.4b} \quad \blacksquare$$

A useful identity concerning variances is that for any constants a and b,

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \qquad (4.6.2)$$

To prove Equation 4.6.2, let $\mu = E[X]$ and recall that $E[aX + b] = a\mu + b$. Thus, by the definition of variance, we have

$$\text{Var}(aX + b) = E[(aX + b - E[aX + b])^2]$$
$$= E[(aX + b - a\mu - b)^2]$$
$$= E[(aX - a\mu)^2]$$
$$= E[a^2(X - \mu)^2]$$
$$= a^2 E[(X - \mu)^2]$$
$$= a^2 \text{Var}(X)$$

Specifying particular values for a and b in Equation 4.6.2 leads to some interesting corollaries. For instance, by setting $a = 0$ in Equation 4.6.2 we obtain that

$$\text{Var}(b) = 0$$

That is, the variance of a constant is 0. (Is this intuitive?) Similarly, by setting $a = 1$ we obtain

$$\text{Var}(X + b) = \text{Var}(X)$$

That is, the variance of a constant plus a random variable is equal to the variance of the random variable. (Is this intuitive? Think about it.) Finally, setting $b = 0$ yields

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

The quantity $\sqrt{\text{Var}(X)}$ is called the *standard deviation* of X. The standard deviation has the same units as does the mean.

REMARK

Analogous to the mean's being the center of gravity of a distribution of mass, the variance represents, in the terminology of mechanics, the moment of inertia.

4.7 COVARIANCE AND VARIANCE OF SUMS OF RANDOM VARIABLES

We showed in Section 4.5 that the expectation of a sum of random variables is equal to the sum of their expectations. The corresponding result for variances is, however, not generally valid. Consider

$$\begin{aligned}
\text{Var}(X + X) &= \text{Var}(2X) \\
&= 2^2 \, \text{Var}(X) \\
&= 4 \, \text{Var}(X) \\
&\neq \text{Var}(X) + \text{Var}(X)
\end{aligned}$$

There is, however, an important case in which the variance of a sum of random variables is equal to the sum of the variances; and this is when the random variables are independent. Before proving this, however, let us define the concept of the covariance of two random variables.

Definition

The *covariance* of two random variables X and Y, written $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

where μ_x and μ_y are the means of X and Y, respectively.

A useful expression for $\text{Cov}(X, Y)$ can be obtained by expanding the right side of the definition. This yields

$$
\begin{aligned}
\text{Cov}(X, Y) &= E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\
&= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y \\
&= E[XY] - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\
&= E[XY] - E[X]E[Y] \tag{4.7.1}
\end{aligned}
$$

From its definition we see that covariance satisfies the following properties:

$$
\text{Cov}(X, Y) = \text{Cov}(Y, X) \tag{4.7.2}
$$

and

$$
\text{Cov}(X, X) = \text{Var}(X) \tag{4.7.3}
$$

Another property of covariance, which immediately follows from its definition, is that, for any constant a,

$$
\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y) \tag{4.7.4}
$$

The proof of Equation 4.7.4 is left as an exercise.

Covariance, like expectation, possesses an additive property.

Lemma 4.7.1

$$
\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)
$$

Proof

$$
\begin{aligned}
\text{Cov}(X_1 &+ X_2, Y) \\
&= E[(X_1 + X_2)Y] - E[X_1 + X_2]E[Y] \quad \text{from Equation 4.7.1} \\
&= E[X_1 Y] + E[X_2 Y] - (E[X_1] + E[X_2])E[Y] \\
&= E[X_1 Y] - E[X_1]E[Y] + E[X_2 Y] - E[X_2]E[Y] \\
&= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y) \quad \blacksquare
\end{aligned}
$$

Lemma 4.7.1 can be easily generalized (see Problem 48) to show that

$$
\text{Cov}\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} \text{Cov}(X_i, Y) \tag{4.7.5}
$$

which gives rise to the following.

PROPOSITION 4.7.2

$$\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\text{Cov}(X_i, Y_j)$$

Proof

$$\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right)$$

$$= \sum_{i=1}^{n}\text{Cov}\left(X_i, \sum_{j=1}^{m} Y_j\right) \quad \text{from Equation 4.7.5}$$

$$= \sum_{i=1}^{n}\text{Cov}\left(\sum_{j=1}^{m} Y_j, X_i\right) \quad \text{by the symmetry property Equation 4.7.2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\text{Cov}(Y_j, X_i) \quad \text{again from Equation 4.7.5}$$

and the result now follows by again applying the symmetry property Equation 4.7.2. ■

Using Equation 4.7.3 gives rise to the following formula for the variance of a sum of random variables.

Corollary 4.7.3

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\text{Var}(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\text{Cov}(X_i, X_j)$$

Proof

Because $\text{Cov}(X, X) = \text{Var}(X)$, for any random variable X, we obtain from Proposition 4.7.2 that

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \left[\sum_{j \neq i} \text{Cov}(X_i, \ X_j) + \text{Cov}(X_i, \ X_i) \right]$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \text{Cov}(X_i, \ X_j) + \sum_{i=1}^{n} \text{Cov}(X_i, \ X_i)$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \text{Cov}(X_i, \ X_j) + \sum_{i=1}^{n} \text{Var}(X_i) \qquad \blacksquare$$

In the case of $n = 2$, Corollary 4.7.3 yields that

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + \text{Cov}(X, Y) + \text{Cov}(Y, X)$$

or, using Equation 4.7.2,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y) \qquad (4.7.6)$$

Theorem 4.7.4

If X and Y are independent random variables, then

$$\text{Cov}(X, \ Y) = 0$$

and so for independent $X_1, \ldots, X_n$,

$$\text{Var}\left(\sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

Proof

We need to prove that $E[XY] = E[X]E[Y]$. Now, in the discrete case,

$$E[XY] = \sum_{j} \sum_{i} x_i y_j P\{X = x_i, Y = y_j\}$$

$$= \sum_{j} \sum_{i} x_i y_j P\{X = x_i\} P\{Y = y_j\} \quad \text{by independence}$$

$$= \sum_{y} y_j P\{Y = y_j\} \sum_{i} x_i P\{X = x_i\}$$

$$= E[Y]E[X]$$

Because a similar argument holds in all other cases, the result is proven. $\blacksquare$

EXAMPLE 4.7a Compute the variance of the sum obtained when 10 independent rolls of a fair die are made.

SOLUTION Letting X_i denote the outcome of the ith roll, we have that

$$\text{Var}\left(\sum_1^{10} X_i\right) = \sum_1^{10} \text{Var}(X_i)$$

$$= 10\tfrac{35}{12} \quad \text{from Example 4.6a}$$

$$= \tfrac{175}{6} \quad \blacksquare$$

EXAMPLE 4.7b Compute the variance of the number of heads resulting from 10 independent tosses of a fair coin.

SOLUTION Letting

$$I_j = \begin{cases} 1 & \text{if the } j\text{th toss lands heads} \\ 0 & \text{if the } j\text{th toss lands tails} \end{cases}$$

then the total number of heads is equal to

$$\sum_{j=1}^{10} I_j$$

Hence, from Theorem 4.7.4,

$$\text{Var}\left(\sum_{j=1}^{10} I_j\right) = \sum_{j=1}^{10} \text{Var}(I_j)$$

Now, since I_j is an indicator random variable for an event having probability $\tfrac{1}{2}$, it follows from Example 4.6b that

$$\text{Var}(I_j) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

and thus

$$\text{Var}\left(\sum_{j=1}^{10} I_j\right) = \frac{10}{4} \quad \blacksquare$$

The covariance of two random variables is important as an indicator of the relationship between them. For instance, consider the situation where X and Y are indicator variables for whether or not the events A and B occur. That is, for events A and B, define

$$X = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}, \qquad Y = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

and note that

$$XY = \begin{cases} 1 & \text{if } X = 1, Y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= P\{X = 1, Y = 1\} - P\{X = 1\}P\{Y = 1\} \end{aligned}$$

From this we see that

$$\begin{aligned} \text{Cov}(X, Y) > 0 &\Leftrightarrow P\{X = 1, Y = 1\} > P\{X = 1\}P\{Y = 1\} \\ &\Leftrightarrow \frac{P\{X = 1, Y = 1\}}{P\{X = 1\}} > P\{Y = 1\} \\ &\Leftrightarrow P\{Y = 1 | X = 1\} > P\{Y = 1\} \end{aligned}$$

that $Y = 1$; whereas the covariance of X and Y is negative if the outcome $X = 1$ makes it less likely that $Y = 1$, and so makes it more likely that $Y = 0$. (By the symmetry of the covariance, the preceding remains true when X and Y are interchanged.)

In general, it can be shown that a positive value of $\text{Cov}(X, Y)$ is an indication that Y tends to increase as X does, whereas a negative value indicates that Y tends to decrease as X increases. The strength of the relationship between X and Y is indicated by the correlation between X and Y, a dimensionless quantity obtained by dividing the covariance by the product of the standard deviations of X and Y. That is,

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

It can be shown (see Problem 49) that this quantity always has a value between -1 and $+1$.

4.8 MOMENT GENERATING FUNCTIONS

The moment generating function $\phi(t)$ of the random variable X is defined for all values t by

$$\phi(t) = E[e^{tX}] = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\phi'(t) = \frac{d}{dt} E[e^{tX}]$$
$$= E\left[\frac{d}{dt}(e^{tX})\right]$$
$$= E[Xe^{tX}]$$

Hence,

$$\phi'(0) = E[X]$$

Similarly,

$$\phi''(t) = \frac{d}{dt}\phi'(t)$$
$$= \frac{d}{dt} E[Xe^{tX}]$$
$$= E\left[\frac{d}{dt}(Xe^{tX})\right]$$
$$= E[X^2 e^{tX}]$$

and so

$$\phi''(0) = E[X^2]$$

In general, the nth derivative of $\phi(t)$ evaluated at $t = 0$ equals $E[X^n]$; that is,

$$\phi^n(0) = E[X^n], \quad n \geq 1$$

An important property of moment generating functions is that the *moment generating function of the sum of independent random variables is just the product of the individual moment generating functions*. To see this, suppose that X and Y are independent and have

moment generating functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then $\phi_{X+Y}(t)$, the moment generating function of $X + Y$, is given by

$$\phi_{X+Y}(t) = E[e^{t(X+Y)}]$$
$$= E[e^{tX}e^{tY}]$$
$$= E[e^{tX}]E[e^{tY}]$$
$$= \phi_X(t)\phi_Y(t)$$

where the next to the last equality follows from Theorem 4.7.4 since X and Y, and thus e^{tX} and e^{tY}, are independent.

Another important result is that the *moment generating function uniquely determines the distribution*. That is, there exists a one-to-one correspondence between the moment generating function and the distribution function of a random variable.

4.9 CHEBYSHEV'S INEQUALITY AND THE WEAK LAW OF LARGE NUMBERS

We start this section by proving a result known as Markov's inequality.

PROPOSITION 4.9.1 MARKOV'S INEQUALITY

If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof

We give a proof for the case where X is continuous with density f.

$$E[X] = \int_0^\infty x f(x)\, dx$$
$$= \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx$$
$$\geq \int_a^\infty x f(x)\, dx$$
$$\geq \int_a^\infty a f(x)\, dx$$
$$= a \int_a^\infty f(x)\, dx$$
$$= a P\{X \geq a\}$$

and the result is proved. ■

As a corollary, we obtain Proposition 4.9.2.

PROPOSITION 4.9.2 CHEBYSHEV'S INEQUALITY

If X is a random variable with mean μ and variance σ^2, then for any value $k > 0$

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

Proof

Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2} \tag{4.9.1}$$

But since $(X - \mu) \geq k^2$ if and only if $|X - \mu| \geq k$, Equation 4.9.1 is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete. ■

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.

EXAMPLE 4.9a Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

   (a) What can be said about the probability that this week's production will exceed 75?

   (b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

SOLUTION Let X be the number of items that will be produced in a week:

   (a) By Markov's inequality

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

   (b) By Chebyshev's inequality

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Hence

$$P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

and so the probability that this week's production will be between 40 and 60 is at least .75.   ■

By replacing k by $k\sigma$ in Equation 4.9.1, we can write Chebyshev's inequality as

$$P\{|X - \mu| > k\sigma\} \leq 1/k^2$$

Thus it states that the probability a random variable differs from its mean by more than k standard deviations is bounded by $1/k^2$.

We will end this section by using Chebyshev's inequality to prove the weak law of large numbers, which states that the probability that the average of the first n terms in a sequence of independent and identically distributed random variables differs by its mean by more than ε goes to 0 as n goes to infinity.

Theorem 4.9.3 The Weak Law of Large Numbers

Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed random variables, each having mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty$$

Proof

We shall prove the result only under the additional assumption that the random variables have a finite variance σ^2. Now, as

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \text{and} \quad \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

it follows from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

and the result is proved.   ■

For an application of the above, suppose that a sequence of independent trials is performed. Let E be a fixed event and denote by $P(E)$ the probability that E occurs on a given trial. Letting

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{if } E \text{ does not occur on trial } i \end{cases}$$

it follows that $X_1 + X_2 + \cdots + X_n$ represents the number of times that E occurs in the first n trials. Because $E[X_i] = P(E)$, it thus follows from the weak law of large numbers that for any positive number ε, no matter how small, the probability that the proportion of the first n trials in which E occurs differs from $P(E)$ by more than ε goes to 0 as n increases.

Problems

1. Five men and 5 women are ranked according to their scores on an examination. Assume that no two scores are alike and all 10! possible rankings are equally likely. Let X denote the highest ranking achieved by a woman (for instance, $X = 2$ if the top-ranked person was male and the next-ranked person was female). Find $P\{X = i\}, i = 1, 2, 3, \ldots, 8, 9, 10$.

2. Let X represent the difference between the number of heads and the number of tails obtained when a coin is tossed n times. What are the possible values of X?

3. In Problem 2, if the coin is assumed fair, for $n = 3$, what are the probabilities associated with the values that X can take on?

4. The distribution function of the random variable X is given

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \le x < 1 \\ \dfrac{2}{3} & 1 \le x < 2 \\ \dfrac{11}{12} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

   (a) Plot this distribution function.
   (b) What is $P\{X > \frac{1}{2}\}$?
   (c) What is $P\{2 < X \le 4\}$?
   (d) What is $P\{X < 3\}$?
   (e) What is $P\{X = 1\}$?

5. Suppose the random variable X has probability density function

$$f(x) = \begin{cases} cx^3, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) Find the value of c.
   (b) Find $P\{.4 < X < .8\}$.

6. The amount of time, in hours, that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that a computer will function between 50 and 150 hours before breaking down? What is the probability that it will function less than 100 hours?

7. The lifetime in hours of a certain kind of radio tube is a random variable having a probability density function given by

$$f(x) = \begin{cases} 0 & x \leq 100 \\ \dfrac{100}{x^2} & x > 100 \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation? Assume that the events $E_i, i = 1, 2, 3, 4, 5$, that the ith such tube will have to be replaced within this time are independent.

8. If the density function of X equals

$$f(x) = \begin{cases} c\, e^{-2x} & 0 < x < \infty \\ 0 & x < 0 \end{cases}$$

find c. What is $P\{X > 2\}$?

9. A set of five transistors are to be tested, one at a time in a random order, to see which of them are defective. Suppose that three of the five transistors are defective, and let N_1 denote the number of tests made until the first defective is spotted, and let N_2 denote the number of additional tests until the second defective is spotted. Find the joint probability mass function of N_1 and N_2.

10. The joint probability density function of X and Y is given by

$$f(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 < x < 1, \quad 0 < y < 2$$

   (a) Verify that this is indeed a joint density function.
   (b) Compute the density function of X.
   (c) Find $P\{X > Y\}$.

11. Let $X_1, X_2, \ldots, X_n$ be independent random variables, each having a uniform distribution over (0, 1). Let $M = \text{maximum } (X_1, X_2, \ldots, X_n)$. Show that the distribution function of M is given by

$$F_M(x) = x^n, \quad 0 \leq x \leq 1$$

What is the probability density function of M?

12. The joint density of X and Y is given by

$$f(x, y) = \begin{cases} xe^{-(x+y)} & x > 0, \, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the density of X.
(b) Compute the density of Y.
(c) Are X and Y independent?

13. The joint density of X and Y is

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

(a) Compute the density of X.
(b) Compute the density of Y.
(c) Are X and Y independent?

14. If the joint density function of X and Y factors into one part depending only on x and one depending only on y, show that X and Y are independent. That is, if

$$f(x, y) = k(x)\, h(y), \quad -\infty < x < \infty, \quad -\infty < y < \infty$$

show that X and Y are independent.

15. Is Problem 14 consistent with the results of Problems 12 and 13?

16. Suppose that X and Y are independent continuous random variables. Show that

(a) $P\{X + Y \le a\} = \displaystyle\int_{-\infty}^{\infty} F_X(a - y) f_Y(y)\, dy$

(b) $P\{X \le Y\} = \displaystyle\int_{-\infty}^{\infty} F_X(y) f_Y(y)\, dy$

where f_Y is the density function of Y, and F_X is the distribution function of X.

17. When a current I (measured in amperes) flows through a resistance R (measured in ohms), the power generated (measured in watts) is given by $W = I^2 R$. Suppose that I and R are independent random variables with densities

$$f_I(x) = 6x(1 - x) \quad 0 \le x \le 1$$
$$f_R(x) = 2x \quad 0 \le x \le 1$$

Determine the density function of W.

18. In Example 4.3b, determine the conditional probability mass function of the size of a randomly chosen family containing 2 girls.

19. Compute the conditional density function of X given $Y = y$ in **(a)** Problem 10 and **(b)** Problem 13.

20. Show that X and Y are independent if and only if

   (a) $p_{X|Y}(x|y) = p_X(x)$ in the discrete case

   (b) $f_{X|Y}(x|y) = f_X(x)$ in the continuous case

21. Compute the expected value of the random variable in Problem 1.

22. Compute the expected value of the random variable in Problem 3.

23. Each night different meteorologists give us the "probability" that it will rain the next day. To judge how well these people predict, we will score each of them as follows: If a meteorologist says that it will rain with probability p, then he or she will receive a score of

$$
\begin{array}{ll}
1 - (1 - p)^2 & \text{if it does rain} \\
1 - p^2 & \text{if it does not rain}
\end{array}
$$

   We will then keep track of scores over a certain time span and conclude that the meteorologist with the highest average score is the best predictor of weather. Suppose now that a given meteorologist is aware of this and so wants to maximize his or her expected score. If this individual truly believes that it will rain tomorrow with probability p^*, what value of p should he or she assert so as to maximize the expected score?

24. An insurance company writes a policy to the effect that an amount of money A must be paid if some event E occurs within a year. If the company estimates that E will occur within a year with probability p, what should it charge the customer so that its expected profit will be 10 percent of A?

25. A total of 4 buses carrying 148 students from the same school arrive at a football stadium. The buses carry, respectively, 40, 33, 25, and 50 students. One of the students is randomly selected. Let X denote the number of students that were on the bus carrying this randomly selected student. One of the 4 bus drivers is also randomly selected. Let Y denote the number of students on her bus.

   (a) Which of $E[X]$ or $E[Y]$ do you think is larger? Why?

   (b) Compute $E[X]$ and $E[Y]$.

26. Suppose that two teams play a series of games that end when one of them has won i games. Suppose that each game played is, independently, won by team A with probability p. Find the expected number of games that are played when $i = 2$. Also show that this number is maximized when $p = \frac{1}{2}$.

27. The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

If $E[X] = \frac{3}{5}$, find a, b.

28. The lifetime in hours of electronic tubes is a random variable having a probability density function given by

$$f(x) = a^2 x e^{-ax}, \quad x \ge 0$$

Compute the expected lifetime of such a tube.

29. Let $X_1, X_2, \ldots, X_n$ be independent random variables having the common density function

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

Find **(a)** $E[\text{Max}(X_1, \ldots, X_n)]$ and **(b)** $E[\text{Min}(X_1, \ldots, X_n)]$.

30. Suppose that X has density function

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

Compute $E[X^n]$ **(a)** by computing the density of X^n and then using the definition of expectation and **(b)** by using Proposition 4.5.1.

31. The time it takes to repair a personal computer is a random variable whose density, in hours, is given by

$$f(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

The cost of the repair depends on the time it takes and is equal to $40 + 30\sqrt{x}$ when the time is x. Compute the expected cost to repair a personal computer.

32. If $E[X] = 2$ and $E[X^2] = 8$, calculate **(a)** $E[(2+4X)^2]$ and **(b)** $E[X^2+(X+1)^2]$.

33. Ten balls are randomly chosen from an urn containing 17 white and 23 black balls. Let X denote the number of white balls chosen. Compute $E[X]$

   (a) by defining appropriate indicator variables X_i, $i = 1, \ldots, 10$ so that

$$X = \sum_{i=1}^{10} X_i$$

(b) by defining appropriate indicator variables $Y_i = 1, \ldots, 17$ so that

$$X = \sum_{i=1}^{17} Y_i$$

34. If X is a continuous random variable having distribution function F, then its *median* is defined as that value of m for which

$$F(m) = 1/2$$

   Find the median of the random variables with density function

   (a) $f(x) = e^{-x}, \quad x \geq 0$;

   (b) $f(x) = 1, \quad 0 \leq x \leq 1$.

35. The median, like the mean, is important in predicting the value of a random variable. Whereas it was shown in the text that the mean of a random variable is the best predictor from the point of view of minimizing the expected value of the square of the error, the median is the best predictor if one wants to minimize the expected value of the absolute error. That is, $E[|X - c|]$ is minimized when c is the median of the distribution function of X. Prove this result when X is continuous with distribution function F and density function f. *Hint:* Write

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

$$= \int_{-\infty}^{c} |x - c| f(x)\, dx + \int_{c}^{\infty} |x - c| f(x)\, dx$$

$$= \int_{-\infty}^{c} (c - x) f(x)\, dx + \int_{c}^{\infty} (x - c) f(x)\, dx$$

$$= c F(c) - \int_{-\infty}^{c} x f(x)\, dx + \int_{c}^{\infty} x f(x)\, dx - c[1 - F(c)]$$

   Now, use calculus to find the minimizing value of c.

36. We say that m_p is the *100p percentile* of the distribution function F if

$$F(m_p) = p$$

   Find m_p for the distribution having density function

$$f(x) = 2e^{-2x}, \quad x \geq 0$$

37. A community consists of 100 married couples. If 50 members of the community die, what is the expected number of marriages that remain intact? Assume that the

set of people who die is equally likely to be any of the $\binom{200}{50}$ groups of size 50. *Hint:* For $i = 1, \ldots, 100$ let

$$X_i = \begin{cases} 1 & \text{if neither member of couple } i \text{ dies} \\ 0 & \text{otherwise} \end{cases}$$

38. Compute the expectation and variance of the number of successes in n independent trials, each of which results in a success with probability p. Is independence necessary?

39. Suppose that X is equally likely to take on any of the values 1, 2, 3, 4. Compute **(a)** $E[X]$ and **(b)** $\text{Var}(X)$.

40. Let $p_i = P\{X = i\}$ and suppose that $p_1 + p_2 + p_3 = 1$. If $E[X] = 2$, what values of p_1, p_2, p_3 **(a)** maximize and **(b)** minimize $\text{Var}(X)$?

41. Compute the mean and variance of the number of heads that appear in 3 flips of a fair coin.

42. Argue that for any random variable X

$$E[X^2] \geq (E[X])^2$$

When does one have equality?

43. A random variable X, which represents the weight (in ounces) of an article, has density function,

$$f(z) = \begin{cases} z - 8 & \text{for } 8 \leq z \leq 9 \\ 10 - z & \text{for } 9 < z \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Calculate the mean and variance of the random variable X.

   (b) The manufacturer sells the article for a fixed price of \$2.00. He guarantees to refund the purchase money to any customer who finds the weight of his article to be less than 8.25 oz. His cost of production is related to the weight of the article by the relation $x/15 + .35$. Find the expected profit per article.

44. Let X_i denote the percentage of votes cast in a given election that are for candidate i, and suppose that X_1 and X_2 have a joint density function

$$f_{X_1,X_2}(x, y) = \begin{cases} 3(x + y), & \text{if } x \geq 0, y \geq 0, 0 \leq x + y \leq 1 \\ 0, & \text{if otherwise} \end{cases}$$

   (a) Find the marginal densities of X_1 and X_2.

   (b) Find $E[X_i]$ and $\text{Var}(X_i)$ for $i = 1, 2$.

45. A product is classified according to the number of defects it contains and the factory that produces it. Let X_1 and X_2 be the random variables that represent the number of defects per unit (taking on possible values of 0, 1, 2, or 3) and the factory number (taking on possible values 1 or 2), respectively. The entries in the table represent the joint possibility mass function of a randomly chosen product.

X_1 \ X_2	1	2
0	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{16}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{4}$

   (a) Find the marginal probability distributions of X_1 and X_2.
   (b) Find $E[X_1]$, $E[X_2]$, $\text{Var}(X_1)$, $\text{Var}(X_2)$, and $\text{Cov}(X_1, X_2)$.

46. Find $\text{Corr}(X_1, X_2)$ for the random variables of Problem 44.
47. Verify Equation 4.7.4.
48. Prove Equation 4.7.5 by using mathematical induction.
49. Let X have variance σ_x^2 and let Y have variance σ_y^2. Starting with

$$0 \leq \text{Var}(X/\sigma_x + Y/\sigma_y)$$

show that

$$-1 \leq \text{Corr}(X, Y)$$

Now using that

$$0 \leq \text{Var}(X/\sigma_x - Y/\sigma_y)$$

conclude that

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Using the result that $\text{Var}(Z) = 0$ implies that Z is constant, argue that, if $\text{Corr}(X, Y) = 1$ or -1, then X and Y are related by

$$Y = a + bx$$

where the sign of b is positive when the correlation is 1 and negative when it is -1.

50. Consider n independent trials, each of which results in any of the outcomes i, $i = 1, 2, 3$, with respective probabilities p_1, p_2, p_3, $\sum_{i=1}^{3} p_i = 1$. Let N_i denote the number of trials that result in outcome i, and show that $\text{Cov}(N_1, N_2) = -np_1p_2$. Also explain why it is intuitive that this covariance is negative. (*Hint:* For $i = 1, \ldots, n$, let

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ results in outcome 1} \\ 0 & \text{if trial } i \text{ does not result in outcome 1} \end{cases}$$

Similarly, for $j = 1, \ldots, n$, let

$$Y_j = \begin{cases} 1 & \text{if trial } j \text{ results in outcome 2} \\ 0 & \text{if trial } j \text{ does not result in outcome 2} \end{cases}$$

Argue that

$$N_1 = \sum_{i=1}^{n} X_i, \quad N_2 = \sum_{j=1}^{n} Y_j$$

Then use Proposition 4.7.2 and Theorem 4.7.4.)

51. In Example 4.5f, compute $\text{Cov}(X_i, X_j)$ and use this result to show that $\text{Var}(X) = 1$.

52. If X_1 and X_2 have the same probability distribution function, show that

$$\text{Cov}(X_1 - X_2, X_1 + X_2) = 0$$

Note that independence is not being assumed.

53. Suppose that X has density function

$$f(x) = e^{-x}, \quad x > 0$$

Compute the moment generating function of X and use your result to determine its mean and variance. Check your answer for the mean by a direct calculation.

54. If the density function of X is

$$f(x) = 1, \quad 0 < x < 1$$

determine $E[e^{tX}]$. Differentiate to obtain $E[X^n]$ and then check your answer.

55. Suppose that X is a random variable with mean and variance both equal to 20. What can be said about $P\{0 \leq X \leq 40\}$?

56. From past experience, a professor knows that the test score of a student taking her final examination is a random variable with mean 75.

   (a) Give an upper bound to the probability that a student's test score will exceed 85.
      Suppose in addition the professor knows that the variance of a student's test score is equal to 25.
   (b) What can be said about the probability that a student will score between 65 and 85?
   (c) How many students would have to take the examination so as to ensure, with probability at least .9, that the class average would be within 5 of 75?

57. Let X and Y have respective distribution functions F_X and F_Y, and suppose that for some constants a and $b > 0$,

$$F_X(x) = F_Y\left(\frac{x-a}{b}\right)$$

   (a) Determine $E[X]$ in terms of $E[Y]$.
   (b) Determine $\text{Var}(X)$ in terms of $\text{Var}(Y)$.

   Hint: X has the same distribution as what other random variable?