Chapter 4

Mathematical Expectation

4.1 Mean of a Random Variable

In Chapter 1, we discussed the sample mean, which is the arithmetic mean of the data. Now consider the following. If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06.$$

This is an average value of the data and yet it is not a possible outcome of $\{0, 1, 2\}$. Hence, an average is not necessarily a possible outcome for the experiment. For instance, a salesman's average monthly income is not likely to be equal to any of his monthly paychecks.

Let us now restructure our computation for the average number of heads so as to have the following equivalent form:

$$(0)\left(\frac{4}{16}\right) + (1)\left(\frac{7}{16}\right) + (2)\left(\frac{5}{16}\right) = 1.06.$$

The numbers 4/16, 7/16, and 5/16 are the fractions of the total tosses resulting in 0, 1, and 2 heads, respectively. These fractions are also the relative frequencies for the different values of X in our experiment. In fact, then, we can calculate the mean, or average, of a set of data by knowing the distinct values that occur and their relative frequencies, without any knowledge of the total number of observations in our set of data. Therefore, if 4/16, or 1/4, of the tosses result in no heads, 7/16 of the tosses result in one head, and 5/16 of the tosses result in two heads, the mean number of heads per toss would be 1.06 no matter whether the total number of tosses were 16, 1000, or even 10,000.

This method of relative frequencies is used to calculate the average number of heads per toss of two coins that we might expect in the long run. We shall refer to this average value as the **mean of the random variable** X or the **mean of the probability distribution of** X and write it as μ_x or simply as μ when it is clear to which random variable we refer. It is also common among statisticians to refer to this mean as the mathematical expectation, or the expected value of the random variable X, and denote it as $E(X)$.

Assuming that 1 fair coin was tossed twice, we find that the sample space for our experiment is

$$S = \{HH, HT, TH, TT\}.$$

Since the 4 sample points are all equally likely, it follows that

$$P(X = 0) = P(TT) = \frac{1}{4}, \quad P(X = 1) = P(TH) + P(HT) = \frac{1}{2},$$

and

$$P(X = 2) = P(HH) = \frac{1}{4},$$

where a typical element, say TH, indicates that the first toss resulted in a tail followed by a head on the second toss. Now, these probabilities are just the relative frequencies for the given events in the long run. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1.$$

This result means that a person who tosses 2 coins over and over again will, on the average, get 1 head per toss.

The method described above for calculating the expected number of heads per toss of 2 coins suggests that the mean, or expected value, of any discrete random variable may be obtained by multiplying each of the values $x_1, x_2, \ldots, x_n$ of the random variable X by its corresponding probability $f(x_1), f(x_2), \ldots, f(x_n)$ and summing the products. This is true, however, only if the random variable is discrete. In the case of continuous random variables, the definition of an expected value is essentially the same with summations replaced by integrations.

Definition 4.1: Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

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

if X is discrete, and

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

if X is continuous.

The reader should note that the way to calculate the expected value, or mean, shown here is different from the way to calculate the sample mean described in Chapter 1, where the sample mean is obtained by using data. In mathematical expectation, the expected value is calculated by using the probability distribution.

However, the mean is usually understood as a "center" value of the underlying distribution if we use the expected value, as in Definition 4.1.

Example 4.1: A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

***Solution*:** Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

Example 4.2: A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn \$1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make \$1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

***Solution*:** First, we know that the salesperson, for the two appointments, can have 4 possible commission totals: \$0, \$1000, \$1500, and \$2500. We then need to calculate their associated probabilities. By independence, we obtain

$$f(\$0) = (1 - 0.7)(1 - 0.4) = 0.18, \quad f(\$2500) = (0.7)(0.4) = 0.28,$$
$$f(\$1000) = (0.7)(1 - 0.4) = 0.42, \text{ and } f(\$1500) = (1 - 0.7)(0.4) = 0.12.$$

Therefore, the expected commission for the salesperson is

$$E(X) = (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) = \$1300.$$

Examples 4.1 and 4.2 are designed to allow the reader to gain some insight into what we mean by the expected value of a random variable. In both cases the random variables are discrete. We follow with an example involving a continuous random variable, where an engineer is interested in the *mean life* of a certain type of electronic device. This is an illustration of a *time to failure* problem that occurs often in practice. The expected value of the life of a device is an important parameter for its evaluation.

Example 4.3: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20{,}000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

***Solution*:** Using Definition 4.1, we have

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20,000}{x^3}\, dx = \int_{100}^{\infty} \frac{20,000}{x^2}\, dx = 200.$$

Therefore, we can expect this type of device to last, *on average*, 200 hours.

Now let us consider a new random variable $g(X)$, which depends on X; that is, each value of $g(X)$ is determined by the value of X. For instance, $g(X)$ might be X^2 or $3X - 1$, and whenever X assumes the value 2, $g(X)$ assumes the value $g(2)$. In particular, if X is a discrete random variable with probability distribution $f(x)$, for $x = -1, 0, 1, 2$, and $g(X) = X^2$, then

$$\begin{aligned} P[g(X) = 0] &= P(X = 0) = f(0), \\ P[g(X) = 1] &= P(X = -1) + P(X = 1) = f(-1) + f(1), \\ P[g(X) = 4] &= P(X = 2) = f(2), \end{aligned}$$

and so the probability distribution of $g(X)$ may be written

$g(x)$	0	1	4
$P[g(X) = g(x)]$	$f(0)$	$f(-1) + f(1)$	$f(2)$

By the definition of the expected value of a random variable, we obtain

$$\begin{aligned} \mu_{g(X)} &= E[g(x)] = 0f(0) + 1[f(-1) + f(1)] + 4f(2) \\ &= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2) = \sum_x g(x) f(x). \end{aligned}$$

This result is generalized in Theorem 4.1 for both discrete and continuous random variables.

Theorem 4.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x) f(x)$$

if X is discrete, and

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

if X is continuous.

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: By Theorem 4.1, the attendant can expect to receive

$$E[g(X)] = E(2X-1) = \sum_{x=4}^{9}(2x-1)f(x)$$
$$= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right)$$
$$+ (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67.$$

Example 4.5: Let X be a random variable with density function

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

Find the expected value of $g(X) = 4X + 3$.

Solution: By Theorem 4.1, we have

$$E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2}(4x^3 + 3x^2)\, dx = 8.$$

We shall now extend our concept of mathematical expectation to the case of two random variables X and Y with joint probability distribution $f(x, y)$.

Definition 4.2: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

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

if X and Y are discrete, and

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

if X and Y are continuous.

Generalization of Definition 4.2 for the calculation of mathematical expectations of functions of several random variables is straightforward.

Example 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X, Y) = XY$. The table is reprinted here for convenience.

		x			Row
	$f(x,y)$	0	1	2	Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

***Solution*:** By Definition 4.2, we write

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y) \\ &= (0)(0)f(0,0) + (0)(1)f(0,1) \\ &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0) \\ &= f(1,1) = \frac{3}{14}. \end{aligned}$$

Example 4.7: Find $E(Y/X)$ for the density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

***Solution*:** We have

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y(1+3y^2)}{4}\ dxdy = \int_0^1 \frac{y+3y^3}{2}\ dy = \frac{5}{8}.$$

Note that if $g(X, Y) = X$ in Definition 4.2, we have

$$E(X) = \begin{cases} \sum\limits_x \sum\limits_y xf(x,y) = \sum\limits_x xg(x) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\ dy\ dx = \int_{-\infty}^{\infty} xg(x)\ dx & \text{(continuous case)}, \end{cases}$$

where $g(x)$ is the marginal distribution of X. Therefore, in calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X. Similarly, we define

$$E(Y) = \begin{cases} \sum\limits_y \sum\limits_x yf(x,y) = \sum\limits_y yh(y) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\ dxdy = \int_{-\infty}^{\infty} yh(y)\ dy & \text{(continuous case)}, \end{cases}$$

where $h(y)$ is the marginal distribution of the random variable Y.

Exercises

4.1 The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given in Exercise 3.13 on page 92 as

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the average number of imperfections per 10 meters of this fabric.

4.2 The probability distribution of the discrete random variable X is

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x}, \quad x = 0, 1, 2, 3.$$

Find the mean of X.

4.3 Find the mean of the random variable T representing the total of the three coins in Exercise 3.25 on page 93.

4.4 A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

4.5 In a gambling game, a woman is paid \$3 if she draws a jack or a queen and \$5 if she draws a king or an ace from an ordinary deck of 52 playing cards. If she draws any other card, she loses. How much should she pay to play if the game is fair?

4.6 An attendant at a car wash is paid according to the number of cars that pass through. Suppose the probabilities are 1/12, 1/12, 1/4, 1/4, 1/6, and 1/6, respectively, that the attendant receives \$7, \$9, \$11, \$13, \$15, or \$17 between 4:00 P.M. and 5:00 P.M. on any sunny Friday. Find the attendant's expected earnings for this particular period.

4.7 By investing in a particular stock, a person can make a profit in one year of \$4000 with probability 0.3 or take a loss of \$1000 with probability 0.7. What is this person's expected gain?

4.8 Suppose that an antique jewelry dealer is interested in purchasing a gold necklace for which the probabilities are 0.22, 0.36, 0.28, and 0.14, respectively, that she will be able to sell it for a profit of \$250, sell it for a profit of \$150, break even, or sell it for a loss of \$150. What is her expected profit?

4.9 A private pilot wishes to insure his airplane for \$200,000. The insurance company estimates that a total loss will occur with probability 0.002, a 50% loss with probability 0.01, and a 25% loss with probability 0.1. Ignoring all other partial losses, what premium should the insurance company charge each year to realize an average profit of \$500?

4.10 Two tire-quality experts examine stacks of tires and assign a quality rating to each tire on a 3-point scale. Let X denote the rating given by expert A and Y denote the rating given by B. The following table gives the joint distribution for X and Y.

	$f(x,y)$	y = 1	2	3
x	1	0.10	0.05	0.02
	2	0.10	0.35	0.05
	3	0.03	0.10	0.20

Find μ_X and μ_Y.

4.11 The density function of coded measurements of the pitch diameter of threads of a fitting is

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

Find the expected value of X.

4.12 If a dealer's profit, in units of \$5000, on a new automobile can be looked upon as a random variable X having the density function

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

find the average profit per automobile.

4.13 The density function of the continuous random variable X, the total number of hours, in units of 100 hours, that a family runs a vacuum cleaner over a period of one year, is given in Exercise 3.7 on page 92 as

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

Find the average number of hours per year that families run their vacuum cleaners.

4.14 Find the proportion X of individuals who can be expected to respond to a certain mail-order solicitation if X has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere}. \end{cases}$$

4.15 Assume that two random variables (X, Y) are uniformly distributed on a circle with radius a. Then the joint probability density function is

$$f(x, y) = \begin{cases} \frac{1}{\pi a^2}, & x^2 + y^2 \le a^2, \\ 0, & \text{otherwise.} \end{cases}$$

Find μ_X, the expected value of X.

4.16 Suppose that you are inspecting a lot of 1000 light bulbs, among which 20 are defectives. You choose two light bulbs randomly from the lot without replacement. Let

$$X_1 = \begin{cases} 1, & \text{if the 1st light bulb is defective,} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if the 2nd light bulb is defective,} \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability that at least one light bulb chosen is defective. [*Hint*: Compute $P(X_1 + X_2 = 1)$.]

4.17 Let X be a random variable with the following probability distribution:

x	-3	6	9
$f(x)$	$1/6$	$1/2$	$1/3$

Find $\mu_{g(X)}$, where $g(X) = (2X + 1)^2$.

4.18 Find the expected value of the random variable $g(X) = X^2$, where X has the probability distribution of Exercise 4.2.

4.19 A large industrial firm purchases several new word processors at the end of each year, the exact number depending on the frequency of repairs in the previous year. Suppose that the number of word processors, X, purchased each year has the following probability distribution:

x	0	1	2	3
$f(x)$	1/10	3/10	2/5	1/5

If the cost of the desired model is \$1200 per unit and at the end of the year a refund of $50X^2$ dollars will be issued, how much can this firm expect to spend on new word processors during this year?

4.20 A continuous random variable X has the density function

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

Find the expected value of $g(X) = e^{2X/3}$.

4.21 What is the dealer's average profit per automobile if the profit on each automobile is given by $g(X) = X^2$, where X is a random variable having the density function of Exercise 4.12?

4.22 The hospitalization period, in days, for patients following treatment for a certain type of kidney disorder is a random variable $Y = X + 4$, where X has the density function

$$f(x) = \begin{cases} \frac{32}{(x+4)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the average number of days that a person is hospitalized following treatment for this disorder.

4.23 Suppose that X and Y have the following joint probability function:

$f(x, y)$		$x = 2$	$x = 4$
y	1	0.10	0.15
	3	0.20	0.30
	5	0.10	0.15

(a) Find the expected value of $g(X, Y) = XY^2$.
(b) Find μ_X and μ_Y.

4.24 Referring to the random variables whose joint probability distribution is given in Exercise 3.39 on page 105,
(a) find $E(X^2Y - 2XY)$;
(b) find $\mu_X - \mu_Y$.

4.25 Referring to the random variables whose joint probability distribution is given in Exercise 3.51 on page 106, find the mean for the total number of jacks and kings when 3 cards are drawn without replacement from the 12 face cards of an ordinary deck of 52 playing cards.

4.26 Let X and Y be random variables with joint density function

$$f(x, y) = \begin{cases} 4xy, & 0 < x, \ y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $Z = \sqrt{X^2 + Y^2}$.

4.27 In Exercise 3.27 on page 93, a density function is given for the time to failure of an important component of a DVD player. Find the mean number of hours to failure of the component and thus the DVD player.

4.28 Consider the information in Exercise 3.28 on page 93. The problem deals with the weight in ounces of the product in a cereal box, with

$$f(x) = \begin{cases} \frac{2}{5}, & 23.75 \le x \le 26.25, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Plot the density function.

(b) Compute the expected value, or mean weight, in ounces.

(c) Are you surprised at your answer in (b)? Explain why or why not.

4.29 Exercise 3.29 on page 93 dealt with an important particle size distribution characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Plot the density function.

(b) Give the mean particle size.

4.30 In Exercise 3.31 on page 94, the distribution of times before a major repair of a washing machine was given as

$$f(y) = \begin{cases} \frac{1}{4}e^{-y/4}, & y \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

What is the population mean of the times to repair?

4.31 Consider Exercise 3.32 on page 94.

(a) What is the mean proportion of the budget allocated to environmental and pollution control?

(b) What is the probability that a company selected at random will have allocated to environmental and pollution control a proportion that exceeds the population mean given in (a)?

4.32 In Exercise 3.13 on page 92, the distribution of the number of imperfections per 10 meters of synthetic fabric is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

(a) Plot the probability function.

(b) Find the expected number of imperfections, $E(X) = \mu$.

(c) Find $E(X^2)$.

4.2 Variance and Covariance of Random Variables

The mean, or expected value, of a random variable X is of special importance in statistics because it describes where the probability distribution is centered. By itself, however, the mean does not give an adequate description of the shape of the distribution. We also need to characterize the variability in the distribution. In Figure 4.1, we have the histograms of two discrete probability distributions that have the same mean, $\mu = 2$, but differ considerably in variability, or the dispersion of their observations about the mean.

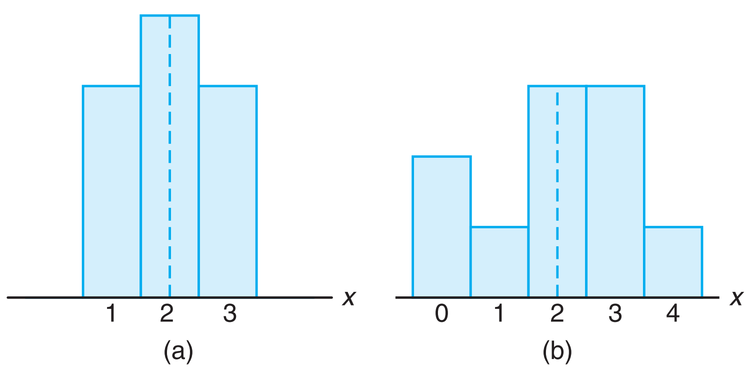

Figure 4.1: Distributions with equal means and unequal dispersions.

The most important measure of variability of a random variable X is obtained by applying Theorem 4.1 with $g(X) = (X - \mu)^2$. The quantity is referred to as the **variance of the random variable** X or the **variance of the probability**

distribution of X and is denoted by $\text{Var}(X)$ or the symbol σ_X^2, or simply by σ^2 when it is clear to which random variable we refer.

Definition 4.3: Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

The quantity $x-\mu$ in Definition 4.3 is called the **deviation of an observation** from its mean. Since the deviations are squared and then averaged, σ^2 will be much smaller for a set of x values that are close to μ than it will be for a set of values that vary considerably from μ.

Example 4.8: Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A [Figure 4.1(a)] is

x	1	2	3
$f(x)$	0.3	0.4	0.3

and that for company B [Figure 4.1(b)] is

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

Solution: For company A, we find that

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

and then

$$\sigma_A^2 = \sum_{x=1}^{3} (x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

For company B, we have

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

and then

$$\begin{aligned}\sigma_B^2 &= \sum_{x=0}^{4} (x-2)^2 f(x) \\ &= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3) \\ &\quad + (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.\end{aligned}$$

Clearly, the variance of the number of automobiles that are used for official business purposes is greater for company B than for company A.

An alternative and preferred formula for finding σ^2, which often simplifies the calculations, is stated in the following theorem.

Theorem 4.2: The variance of a random variable X is

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

***Proof*:** For the discrete case, we can write

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$
$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).$$

Since $\mu = \sum_x x f(x)$ by definition, and $\sum_x f(x) = 1$ for any discrete probability distribution, it follows that

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

For the continuous case the proof is step by step the same, with summations replaced by integrations.

Example 4.9: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2.

***Solution*:** First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Example 4.10: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

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

Find the mean and variance of X.

***Solution*:** Calculating $E(X)$ and $E(X^2)$, we have

$$\mu = E(X) = 2\int_1^2 x(x-1)\ dx = \frac{5}{3}$$

and

$$E(X^2) = 2\int_1^2 x^2(x-1)\ dx = \frac{17}{6}.$$

Therefore,

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

At this point, the variance or standard deviation has meaning only when we compare two or more distributions that have the same units of measurement. Therefore, we could compare the variances of the distributions of contents, measured in liters, of bottles of orange juice from two companies, and the larger value would indicate the company whose product was more variable or less uniform. It would not be meaningful to compare the variance of a distribution of heights to the variance of a distribution of aptitude scores. In Section 4.4, we show how the standard deviation can be used to describe a single distribution of observations.

We shall now extend our concept of the variance of a random variable X to include random variables related to X. For the random variable $g(X)$, the variance is denoted by $\sigma^2_{g(X)}$ and is calculated by means of the following theorem.

Theorem 4.3: Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

if X is discrete, and

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\ dx$$

if X is continuous.

***Proof*:** Since $g(X)$ is itself a random variable with mean $\mu_{g(X)}$ as defined in Theorem 4.1, it follows from Definition 4.3 that

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]\}.$$

Now, applying Theorem 4.1 again to the random variable $[g(X)-\mu_{g(X)}]^2$ completes the proof.

Example 4.11: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

***Solution*:** First, we find the mean of the random variable $2X+3$. According to Theorem 4.1,

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3}(2x+3)f(x) = 6.$$

Now, using Theorem 4.3, we have

$$\begin{aligned}\sigma^2_{2X+3} &= E\{[(2X+3) - \mu_{2x+3}]^2\} = E[(2X+3-6)^2] \\ &= E(4X^2 - 12X + 9) = \sum_{x=0}^{3}(4x^2 - 12x + 9)f(x) = 4.\end{aligned}$$

Example 4.12: Let X be a random variable having the density function given in Example 4.5 on page 115. Find the variance of the random variable $g(X) = 4X + 3$.

***Solution*:** In Example 4.5, we found that $\mu_{4X+3} = 8$. Now, using Theorem 4.3,

$$\begin{aligned}\sigma^2_{4X+3} &= E\{[(4X+3) - 8]^2\} = E[(4X-5)^2] \\ &= \int_{-1}^{2}(4x-5)^2\frac{x^2}{3}\,dx = \frac{1}{3}\int_{-1}^{2}(16x^4 - 40x^3 + 25x^2)\,dx = \frac{51}{5}.\end{aligned}$$

If $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, where $\mu_X = E(X)$ and $\mu_Y = E(Y)$, Definition 4.2 yields an expected value called the **covariance** of X and Y, which we denote by σ_{XY} or $\text{Cov}(X,Y)$.

Definition 4.4: Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x\sum_y(x-\mu_X)(y-\mu_y)f(x,y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_y)f(x,y)\,dx\,dy$$

if X and Y are continuous.

The covariance between two random variables is a measure of the nature of the association between the two. If large values of X often result in large values of Y or small values of X result in small values of Y, positive $X-\mu_X$ will often result in positive $Y-\mu_Y$ and negative $X-\mu_X$ will often result in negative $Y-\mu_Y$. Thus, the product $(X-\mu_X)(Y-\mu_Y)$ will tend to be positive. On the other hand, if large X values often result in small Y values, the product $(X-\mu_X)(Y-\mu_Y)$ will tend to be negative. The *sign* of the covariance indicates whether the relationship between two dependent random variables is positive or negative. When X and Y are statistically independent, it can be shown that the covariance is zero (see Corollary 4.5). The converse, however, is not generally true. Two variables may have zero covariance and still not be statistically independent. Note that the covariance only describes the *linear* relationship between two random variables. Therefore, if a covariance between X and Y is zero, X and Y may have a nonlinear relationship, which means that they are not necessarily independent.

The alternative and preferred formula for σ_{XY} is stated by Theorem 4.4.

Theorem 4.4: The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y.$$

***Proof*:** For the discrete case, we can write

$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y)\\ &= \sum_x\sum_y xyf(x,y) - \mu_X \sum_x\sum_y yf(x,y)\\ &\quad - \mu_Y \sum_x\sum_y xf(x,y) + \mu_X\mu_Y \sum_x\sum_y f(x,y).\end{aligned}$$

Since

$$\mu_X = \sum_x xf(x,y), \quad \mu_Y = \sum_y yf(x,y), \text{ and } \sum_x\sum_y f(x,y) = 1$$

for any joint discrete distribution, it follows that

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y.$$

For the continuous case, the proof is identical with summations replaced by integrals.

Example 4.13: Example 3.14 on page 95 describes a situation involving the number of blue refills X and the number of red refills Y. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

	$f(x,y)$	$x=0$	$x=1$	$x=2$	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y.

***Solution*:** From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

and

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Example 4.14: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Solution: We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4 \, dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1-y^2) \, dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2 \, dx \, dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}.$$

Although the covariance between two random variables does provide information regarding the nature of the relationship, the magnitude of σ_{XY} *does not indicate anything regarding the strength of the relationship*, since σ_{XY} is not scale-free. Its magnitude will depend on the units used to measure both X and Y. There is a scale-free version of the covariance called the **correlation coefficient** that is used widely in statistics.

Definition 4.5: Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

It should be clear to the reader that ρ_{XY} is free of the units of X and Y. The correlation coefficient satisfies the inequality $-1 \le \rho_{XY} \le 1$. It assumes a value of zero when $\sigma_{XY} = 0$. Where there is an exact linear dependency, say $Y \equiv a + bX$,

$\rho_{XY} = 1$ if $b > 0$ and $\rho_{XY} = -1$ if $b < 0$. (See Exercise 4.48.) The correlation coefficient is the subject of more discussion in Chapter 12, where we deal with linear regression.

Example 4.15: Find the correlation coefficient between X and Y in Example 4.13.

Solution: Since

$$E(X^2) = (0^2)\left(\frac{5}{14}\right) + (1^2)\left(\frac{15}{28}\right) + (2^2)\left(\frac{3}{28}\right) = \frac{27}{28}$$

and

$$E(Y^2) = (0^2)\left(\frac{15}{28}\right) + (1^2)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{4}{7},$$

we obtain

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112} \text{ and } \sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

Example 4.16: Find the correlation coefficient of X and Y in Example 4.14.

Solution: Because

$$E(X^2) = \int_0^1 4x^5\, dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2)\, dy = 1 - \frac{2}{3} = \frac{1}{3},$$

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

Note that although the covariance in Example 4.15 is larger in magnitude (disregarding the sign) than that in Example 4.16, the relationship of the magnitudes of the correlation coefficients in these two examples is just the reverse. This is evidence that we cannot look at the magnitude of the covariance to decide on how strong the relationship is.

Exercises

4.33 Use Definition 4.3 on page 120 to find the variance of the random variable X of Exercise 4.7 on page 117.

4.34 Let X be a random variable with the following probability distribution:

x	-2	3	5
$f(x)$	0.3	0.2	0.5

Find the standard deviation of X.

4.35 The random variable X, representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
$f(x)$	0.01	0.25	0.4	0.3	0.04

Using Theorem 4.2 on page 121, find the variance of X.

4.36 Suppose that the probabilities are 0.4, 0.3, 0.2, and 0.1, respectively, that 0, 1, 2, or 3 power failures will strike a certain subdivision in any given year. Find the mean and variance of the random variable X representing the number of power failures striking this subdivision.

4.37 A dealer's profit, in units of \$5000, on a new automobile is a random variable X having the density function given in Exercise 4.12 on page 117. Find the variance of X.

4.38 The proportion of people who respond to a certain mail-order solicitation is a random variable X having the density function given in Exercise 4.14 on page 117. Find the variance of X.

4.39 The total number of hours, in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a random variable X having the density function given in Exercise 4.13 on page 117. Find the variance of X.

4.40 Referring to Exercise 4.14 on page 117, find $\sigma^2_{g(X)}$ for the function $g(X) = 3X^2 + 4$.

4.41 Find the standard deviation of the random variable $g(X) = (2X + 1)^2$ in Exercise 4.17 on page 118.

4.42 Using the results of Exercise 4.21 on page 118, find the variance of $g(X) = X^2$, where X is a random variable having the density function given in Exercise 4.12 on page 117.

4.43 The length of time, in minutes, for an airplane to obtain clearance for takeoff at a certain airport is a random variable $Y = 3X - 2$, where X has the density function

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

Find the mean and variance of the random variable Y.

4.44 Find the covariance of the random variables X and Y of Exercise 3.39 on page 105.

4.45 Find the covariance of the random variables X and Y of Exercise 3.49 on page 106.

4.46 Find the covariance of the random variables X and Y of Exercise 3.44 on page 105.

4.47 For the random variables X and Y whose joint density function is given in Exercise 3.40 on page 105, find the covariance.

4.48 Given a random variable X, with standard deviation σ_X, and a random variable $Y = a + bX$, show that if $b < 0$, the correlation coefficient $\rho_{XY} = -1$, and if $b > 0$, $\rho_{XY} = 1$.

4.49 Consider the situation in Exercise 4.32 on page 119. The distribution of the number of imperfections per 10 meters of synthetic failure is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the variance and standard deviation of the number of imperfections.

4.50 For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

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

Find the variance and standard deviation of X.

4.51 For the random variables X and Y in Exercise 3.39 on page 105, determine the correlation coefficient between X and Y.

4.52 Random variables X and Y follow a joint distribution

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

Determine the correlation coefficient between X and Y.

4.3 Means and Variances of Linear Combinations of Random Variables

We now develop some useful properties that will simplify the calculations of means and variances of random variables that appear in later chapters. These properties will permit us to deal with expectations in terms of other parameters that are either known or easily computed. All the results that we present here are valid for both discrete and continuous random variables. Proofs are given only for the continuous case. We begin with a theorem and two corollaries that should be, intuitively, reasonable to the reader.

Theorem 4.5: If a and b are constants, then

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

***Proof*:** By the definition of expected value,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x)\ dx = a \int_{-\infty}^{\infty} x f(x)\ dx + b \int_{-\infty}^{\infty} f(x)\ dx.$$

The first integral on the right is $E(X)$ and the second integral equals 1. Therefore, we have

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

Corollary 4.1: Setting $a = 0$, we see that $E(b) = b$.

Corollary 4.2: Setting $b = 0$, we see that $E(aX) = aE(X)$.

Example 4.17: Applying Theorem 4.5 to the discrete random variable $f(X) = 2X - 1$, rework Example 4.4 on page 115.

***Solution*:** According to Theorem 4.5, we can write

$$E(2X - 1) = 2E(X) - 1.$$

Now

$$\mu = E(X) = \sum_{x=4}^{9} x f(x)$$
$$= (4)\left(\frac{1}{12}\right) + (5)\left(\frac{1}{12}\right) + (6)\left(\frac{1}{4}\right) + (7)\left(\frac{1}{4}\right) + (8)\left(\frac{1}{6}\right) + (9)\left(\frac{1}{6}\right) = \frac{41}{6}.$$

Therefore,

$$\mu_{2X-1} = (2)\left(\frac{41}{6}\right) - 1 = \$12.67,$$

as before.

Example 4.18: Applying Theorem 4.5 to the continuous random variable $g(X) = 4X + 3$, rework Example 4.5 on page 115.

Solution: For Example 4.5, we may use Theorem 4.5 to write

$$E(4X + 3) = 4E(X) + 3.$$

Now

$$E(X) = \int_{-1}^{2} x\left(\frac{x^2}{3}\right) dx = \int_{-1}^{2} \frac{x^3}{3}\, dx = \frac{5}{4}.$$

Therefore,

$$E(4X + 3) = (4)\left(\frac{5}{4}\right) + 3 = 8,$$

as before.

Theorem 4.6: The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

Proof: By definition,

$$\begin{aligned} E[g(X) \pm h(X)] &= \int_{-\infty}^{\infty} [g(x) \pm h(x)]f(x)\, dx \\ &= \int_{-\infty}^{\infty} g(x)f(x)\, dx \pm \int_{-\infty}^{\infty} h(x)f(x)\, dx \\ &= E[g(X)] \pm E[h(X)]. \end{aligned}$$

Example 4.19: Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X - 1)^2$.

Solution: Applying Theorem 4.6 to the function $Y = (X - 1)^2$, we can write

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

From Corollary 4.1, $E(1) = 1$, and by direct computation,

$$\begin{aligned} E(X) &= (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1 \text{ and} \\ E(X^2) &= (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2. \end{aligned}$$

Hence,

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

Example 4.20: The weekly demand for a certain drink, in thousands of liters, at a chain of convenience stores is a continuous random variable $g(X) = X^2 + X - 2$, where X has the density function

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

Find the expected value of the weekly demand for the drink.

***Solution*:** By Theorem 4.6, we write

$$E(X^2 + X - 2) = E(X^2) + E(X) - E(2).$$

From Corollary 4.1, $E(2) = 2$, and by direct integration,

$$E(X) = \int_1^2 2x(x-1)\ dx = \frac{5}{3} \text{ and } E(X^2) = \int_1^2 2x^2(x-1)\ dx = \frac{17}{6}.$$

Now

$$E(X^2 + X - 2) = \frac{17}{6} + \frac{5}{3} - 2 = \frac{5}{2},$$

so the average weekly demand for the drink from this chain of efficiency stores is 2500 liters.

Suppose that we have two random variables X and Y with joint probability distribution $f(x, y)$. Two additional properties that will be very useful in succeeding chapters involve the expected values of the sum, difference, and product of these two random variables. First, however, let us prove a theorem on the expected value of the sum or difference of functions of the given variables. This, of course, is merely an extension of Theorem 4.6.

Theorem 4.7: The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,

$$E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)].$$

***Proof*:** By Definition 4.2,

$$\begin{aligned} E[g(X, Y) \pm h(X, Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [g(x, y) \pm h(x, y)] f(x, y)\ dx\ dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\ dx\ dy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y) f(x, y)\ dx\ dy \\ &= E[g(X, Y)] \pm E[h(X, Y)]. \end{aligned}$$

Corollary 4.3: Setting $g(X, Y) = g(X)$ and $h(X, Y) = h(Y)$, we see that

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)].$$

Corollary 4.4: Setting $g(X, Y) = X$ and $h(X, Y) = Y$, we see that

$$E[X \pm Y] = E[X] \pm E[Y].$$

If X represents the daily production of some item from machine A and Y the daily production of the same kind of item from machine B, then $X + Y$ represents the total number of items produced daily by both machines. Corollary 4.4 states that the average daily production for both machines is equal to the sum of the average daily production of each machine.

Theorem 4.8: Let X and Y be two independent random variables. Then

$$E(XY) = E(X)E(Y).$$

***Proof*:** By Definition 4.2,

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)\ dx\ dy.$$

Since X and Y are independent, we may write

$$f(x, y) = g(x)h(y),$$

where $g(x)$ and $h(y)$ are the marginal distributions of X and Y, respectively. Hence,

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyg(x)h(y)\ dx\ dy = \int_{-\infty}^{\infty} xg(x)\ dx \int_{-\infty}^{\infty} yh(y)\ dy$$
$$= E(X)E(Y).$$

Theorem 4.8 can be illustrated for discrete variables by considering the experiment of tossing a green die and a red die. Let the random variable X represent the outcome on the green die and the random variable Y represent the outcome on the red die. Then XY represents the product of the numbers that occur on the pair of dice. In the long run, the average of the products of the numbers is equal to the product of the average number that occurs on the green die and the average number that occurs on the red die.

Corollary 4.5: Let X and Y be two independent random variables. Then $\sigma_{XY} = 0$.

***Proof*:** The proof can be carried out by using Theorems 4.4 and 4.8.

Example 4.21: It is known that the ratio of gallium to arsenide does not affect the functioning of gallium-arsenide wafers, which are the main components of microchips. Let X denote the ratio of gallium to arsenide and Y denote the functional wafers retrieved during a 1-hour period. X and Y are independent random variables with the joint density function

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(XY) = E(X)E(Y)$, as Theorem 4.8 suggests.

***Solution*:** By definition,

$$E(XY) = \int_0^1 \int_0^2 \frac{x^2y(1+3y^2)}{4}\, dxdy = \frac{5}{6},\ E(X) = \frac{4}{3},\ \text{and } E(Y) = \frac{5}{8}.$$

Hence,

$$E(X)E(Y) = \left(\frac{4}{3}\right)\left(\frac{5}{8}\right) = \frac{5}{6} = E(XY).$$

We conclude this section by proving one theorem and presenting several corollaries that are useful for calculating variances or standard deviations.

Theorem 4.9: If X and Y are random variables with joint probability distribution $f(x, y)$ and a, b, and c are constants, then

$$\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

***Proof*:** By definition, $\sigma^2_{aX+bY+c} = E\{[(aX + bY + c) - \mu_{aX+bY+c}]^2\}$. Now

$$\mu_{aX+bY+c} = E(aX + bY + c) = aE(X) + bE(Y) + c = a\mu_X + b\mu_Y + c,$$

by using Corollary 4.4 followed by Corollary 4.2. Therefore,

$$\begin{aligned}\sigma^2_{aX+bY+c} &= E\{[a(X-\mu_X) + b(Y-\mu_Y)]^2\}\\ &= a^2E[(X-\mu_X)^2] + b^2E[(Y-\mu_Y)^2] + 2abE[(X-\mu_X)(Y-\mu_Y)]\\ &= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.\end{aligned}$$

Using Theorem 4.9, we have the following corollaries.

Corollary 4.6: Setting $b = 0$, we see that

$$\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollary 4.7: Setting $a = 1$ and $b = 0$, we see that

$$\sigma^2_{X+c} = \sigma^2_X = \sigma^2.$$

Corollary 4.8: Setting $b = 0$ and $c = 0$, we see that

$$\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollaries 4.6 and 4.7 state that the variance is unchanged if a constant is added to or subtracted from a random variable. The addition or subtraction of a constant simply shifts the values of X to the right or to the left but does not change their variability. However, if a random variable is multiplied or divided by a constant, then Corollaries 4.6 and 4.8 state that the variance is multiplied or divided by the square of the constant.

Corollary 4.9: If X and Y are independent random variables, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

The result stated in Corollary 4.9 is obtained from Theorem 4.9 by invoking Corollary 4.5.

Corollary 4.10: If X and Y are independent random variables, then

$$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Corollary 4.10 follows when b in Corollary 4.9 is replaced by $-b$. Generalizing to a linear combination of n independent random variables, we have Corollary 4.11.

Corollary 4.11: If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\sigma^2_{a_1X_1+a_2X_2+\cdots+a_nX_n} = a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + \cdots + a_n^2\sigma^2_{X_n}.$$

Example 4.22: If X and Y are random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

Solution:

$$\begin{aligned}
\sigma^2_Z &= \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y} && \text{(by Corollary 4.6)}\\
&= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma_{XY} && \text{(by Theorem 4.9)}\\
&= (9)(2) + (16)(4) - (24)(-2) = 130.
\end{aligned}$$

Example 4.23: Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

Solution:

$$\begin{aligned}
\sigma^2_Z &= \sigma^2_{3X-2Y+5} = \sigma^2_{3X-2Y} && \text{(by Corollary 4.6)}\\
&= 9\sigma^2_x + 4\sigma^2_y && \text{(by Corollary 4.10)}\\
&= (9)(2) + (4)(3) = 30.
\end{aligned}$$

What If the Function Is Nonlinear?

In that which has preceded this section, we have dealt with properties of linear functions of random variables for very important reasons. Chapters 8 through 15 will discuss and illustrate practical real-world problems in which the analyst is constructing a **linear model** to describe a data set and thus to describe or explain the behavior of a certain scientific phenomenon. Thus, it is natural that expected values and variances of linear combinations of random variables are encountered. However, there are situations in which properties of **nonlinear** functions of random variables become important. Certainly there are many scientific phenomena that are nonlinear, and certainly statistical modeling using nonlinear functions is very important. In fact, in Chapter 12, we deal with the modeling of what have become standard nonlinear models. Indeed, even a simple function of random variables, such as $Z = X/Y$, occurs quite frequently in practice, and yet unlike in the case of the expected value of linear combinations of random variables, there is no simple general rule. For example,

$$E(Z) = E(X/Y) \neq E(X)/E(Y),$$

except in very special circumstances.

The material provided by Theorems 4.5 through 4.9 and the various corollaries is extremely useful in that there are no restrictions on the form of the density or probability functions, apart from the property of independence when it is required as in the corollaries following Theorems 4.9. To illustrate, consider Example 4.23; the variance of $Z = 3X - 2Y + 5$ does not require restrictions on the distributions of the amounts X and Y of the two types of impurities. Only independence between X and Y is required. Now, we do have at our disposal the capacity to find $\mu_{g(X)}$ and $\sigma^2_{g(X)}$ for any function $g(\cdot)$ from first principles established in Theorems 4.1 and 4.3, where it is assumed that the corresponding distribution $f(x)$ is **known**. Exercises 4.40, 4.41, and 4.42, among others, illustrate the use of these theorems. Thus, if the function $g(x)$ is nonlinear and the density function (or probability function in the discrete case) is known, $\mu_{g(X)}$ and $\sigma^2_{g(X)}$ can be evaluated exactly. But, similar to the rules given for linear combinations, are there rules for nonlinear functions that can be used when the form of the distribution of the pertinent random variables is not known?

In general, suppose X is a random variable and $Y = g(x)$. The general solution for $E(Y)$ or $\text{Var}(Y)$ can be difficult to find and depends on the complexity of the function $g(\cdot)$. However, there are approximations available that depend on a linear approximation of the function $g(x)$. For example, suppose we denote $E(X)$ as μ and $\text{Var}(X) = \sigma_X^2$. Then a Taylor series approximation of $g(x)$ around $X = \mu_X$ gives

$$g(x) = g(\mu_X) + \left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_X} (x - \mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \frac{(x-\mu_X)^2}{2} + \cdots .$$

As a result, if we truncate after the linear term and take the expected value of both sides, we obtain $E[g(X)] \approx g(\mu_X)$, which is certainly intuitive and in some cases gives a reasonable approximation. However, if we include the second-order term of the Taylor series, then we have a second-order adjustment for this *first-order approximation* as follows:

Approximation of $E[g(X)]$

$$E[g(X)] \approx g(\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \frac{\sigma_X^2}{2}.$$

Example 4.24: Given the random variable X with mean μ_X and variance σ_X^2, give the second-order approximation to $E(e^X)$.

Solution: Since $\frac{\partial e^x}{\partial x} = e^x$ and $\frac{\partial^2 e^x}{\partial x^2} = e^x$, we obtain $E(e^X) \approx e^{\mu_X}(1 + \sigma_X^2/2)$.

Similarly, we can develop an approximation for $\text{Var}[g(x)]$ by taking the variance of both sides of the first-order Taylor series expansion of $g(x)$.

Approximation of $\text{Var}[g(X)]$

$$\text{Var}[g(X)] \approx \left[\frac{\partial g(x)}{\partial x}\right]^2_{x=\mu_X} \sigma_X^2.$$

Example 4.25: Given the random variable X as in Example 4.24, give an approximate formula for $\text{Var}[g(x)]$.

Solution: Again $\frac{\partial e^x}{\partial x} = e^x$; thus, $\text{Var}(X) \approx e^{2\mu_X}\sigma_X^2$.

These approximations can be extended to nonlinear functions of more than one random variable.

Given a set of independent random variables $X_1, X_2, \ldots, X_k$ with means $\mu_1, \mu_2, \ldots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, respectively, let

$$Y = h(X_1, X_2, \ldots, X_k)$$

be a nonlinear function; then the following are approximations for $E(Y)$ and $\text{Var}(Y)$:

$$E(Y) \approx h(\mu_1, \mu_2, \ldots, \mu_k) + \sum_{i=1}^{k} \frac{\sigma_i^2}{2} \left[\frac{\partial^2 h(x_1, x_2, \ldots, x_k)}{\partial x_i^2} \right] \bigg|_{x_i=\mu_i,\ 1\le i\le k},$$

$$\text{Var}(Y) \approx \sum_{i=1}^{k} \left[\frac{\partial h(x_1, x_2, \ldots, x_k)}{\partial x_i} \right]^2 \bigg|_{x_i=\mu_i,\ 1\le i\le k} \sigma_i^2.$$

Example 4.26: Consider two independent random variables X and Z with means μ_X and μ_Z and variances σ_X^2 and σ_Z^2, respectively. Consider a random variable

$$Y = X/Z.$$

Give approximations for $E(Y)$ and $\text{Var}(Y)$.

Solution: For $E(Y)$, we must use $\frac{\partial y}{\partial x} = \frac{1}{z}$ and $\frac{\partial y}{\partial z} = -\frac{x}{z^2}$. Thus,

$$\frac{\partial^2 y}{\partial x^2} = 0 \text{ and } \frac{\partial^2 y}{\partial z^2} = \frac{2x}{z^3}.$$

As a result,

$$E(Y) \approx \frac{\mu_X}{\mu_Z} + \frac{\mu_X}{\mu_Z^3}\sigma_Z^2 = \frac{\mu_X}{\mu_Z}\left(1 + \frac{\sigma_Z^2}{\mu_Z^2}\right),$$

and the approximation for the variance of Y is given by

$$\text{Var}(Y) \approx \frac{1}{\mu_Z^2}\sigma_X^2 + \frac{\mu_X^2}{\mu_Z^4}\sigma_Z^2 = \frac{1}{\mu_Z^2}\left(\sigma_X^2 + \frac{\mu_X^2}{\mu_Z^2}\sigma_Z^2\right).$$

4.4 Chebyshev's Theorem

In Section 4.2 we stated that the variance of a random variable tells us something about the variability of the observations about the mean. If a random variable has a small variance or standard deviation, we would expect most of the values to be grouped around the mean. Therefore, the probability that the random variable assumes a value within a certain interval about the mean is greater than for a similar random variable with a larger standard deviation. If we think of probability in terms of area, we would expect a continuous distribution with a large value of σ to indicate a greater variability, and therefore we should expect the area to be more spread out, as in Figure 4.2(a). A distribution with a small standard deviation should have most of its area close to μ, as in Figure 4.2(b).