CHAPTER

5

JOINT DISTRIBUTIONS

Thus far interest has centered on a single random variable of either the discrete or the continuous type. Such random variables are called *univariate*. Problems do arise in which two random variables are to be studied simultaneously. For example, we might wish to study the yield of a chemical reaction in conjunction with the temperature at which the reaction is run. Typical questions to ask are: "Is the yield independent of the temperature?" or, "What is the average yield if the temperature is 40° C?" To answer questions of this type, we need to study what are called *two-dimensional* or *bivariate random variables* of both the discrete and continuous type. In this chapter we present a brief introduction to the basic theoretical concepts underlying these variables. These concepts form the basis for the study of regression analysis and correlation, topics of extreme importance in applied statistics. (See Chaps. 11 and 12.)

5.1 JOINT DENSITIES AND INDEPENDENCE

We begin by considering two-dimensional random variables and their density functions. The definitions presented here are natural extensions of those presented for a single random variable in Chaps. 3 and 4. (See Definition 3.2.1 and 4.1.2.)

Definition 5.1.1 (Discrete joint density). Let X and Y be discrete random variables. The ordered pair (X, Y) is called a two-dimensional discrete random variable. A function f_{XY} such that

$$f_{XY}(x, y) = P[X = x \text{ and } Y = y]$$

is called the joint density for (X, Y).

Again, let us point out that in the discrete case some statisticians prefer to use the term "probability function" or "probability mass function" rather than the term "density." We shall use the term "density" and the notation f_{XY} in both the discrete and the continuous cases for consistency of notation and terminology.

Note that the purpose of the density here is the same as in the past—to allow us to compute the probability that the random variable (X, Y) will assume specific values. As in the one-dimensional case, f_{XY} is nonnegative since it represents a probability. Furthermore, if the density is summed over all possible values of X and Y, it must sum to 1. That is, the necessary and sufficient conditions for a function to be a joint density for a two-dimensional discrete random variable are as follows:

Necessary and Sufficient Conditions for a Function to Be a Discrete Joint Density

1. $f_{XY}(x, y) \geq 0$

2. $\sum_{\text{all } x} \sum_{\text{all } y} f_{XY}(x, y) = 1$

The joint density in the discrete case is sometimes expressed in closed form. However, it is more common to present the density in table form.

Example 5.1.1. In an automobile plant two tasks are performed by robots. The first entails welding two joints; the second, tightening three bolts. Let X denote the number of defective welds and Y the number of improperly tightened bolts produced per car. Since X and Y are each discrete, (X, Y) is a two-dimensional discrete random variable. Past data indicates that the joint density for (X, Y) is as shown in Table 5.1. Note that each entry in the table is a number between 0 and 1 and therefore can be interpreted as a probability. Furthermore,

$$\sum_{x=0}^{2} \sum_{y=0}^{3} f_{XY}(x, y) = .840 + .030 + .020 + \cdots + .001 = 1$$

as required. The probability that there will be no errors made by the robots is given by

$$P[X = 0 \text{ and } Y = 0] = f_{XY}(0, 0) = .840$$

The probability that there will be exactly one error made is

TABLE 5.1

x/y	0	1	2	3
0	.840	.030	.020	.010
1	.060	.010	.008	.002
2	.010	.005	.004	.001

$$\begin{aligned} P[X = 1 \text{ and } Y = 0] + P[X = 0 \text{ and } Y = 1] &= f_{XY}(1, 0) + f_{XY}(0, 1) \\ &= .060 + .030 \\ &= .09 \end{aligned}$$

The probability that there will be no improperly tightened bolts is $P[Y = 0]$. Note that this probability, which concerns only the random variable Y, can be obtained by summing $f_{XY}(x, 0)$ over all values of X. That is,

$$\begin{aligned} P[Y = 0] &= \sum_{x=0}^{2} f_{XY}(x, 0) \\ &= P[X = 0 \text{ and } Y = 0] + P[X = 1 \text{ and } Y = 0] \\ &\quad + P[X = 2 \text{ and } Y = 0] \\ &= .840 + .060 + .010 = .91 \end{aligned}$$

Marginal Distributions: Discrete

Given the joint density for a two-dimensional discrete random variable (X, Y), it is easy to derive the individual densities for X and Y. The manner in which this is done is suggested by the method used to answer the last question posed in Example 5.1.1. To find the density for Y alone, we sum the joint density over all values of X; to find the density for X alone, we sum over Y. When the joint density is given in table form, it is customary to report the individual densities for X and Y in the margins of the joint density table. For this reason, the densities for X and Y alone are called *marginal densities*. This idea is formalized in Definition 5.1.2.

Definition 5.1.2 (Discrete marginal densities). Let (X, Y) be a two-dimensional discrete random variable with joint density f_{XY}. The marginal density for X, denoted by f_X, is given by

$$f_X(x) = \sum_{\text{all } y} f_{XY}(x, y)$$

The marginal density for Y, denoted by f_Y, is given by

$$f_Y(y) = \sum_{\text{all } x} f_{XY}(x, y)$$

Example 5.1.2. Table 5.2 gives the joint density for the random variable (X, Y) of Example 5.1.1. It also displays the marginal densities for X, the number of defective welds, and Y, the number of improperly tightened bolts per car. Note that the marginal density for X is obtained by summing across the rows of the table; that for Y is obtained by summing down the columns.

Joint and Marginal Distributions: Continuous

The idea of a two-dimensional continuous random variable and continuous joint density can be developed by extending Definition 4.1.1 to more than one variable.

TABLE 5.2

x/y	0	1	2	3	$f_X(x)$
0	.840	.030	.020	.010	.900
1	.060	.010	.008	.002	.080
2	.010	.005	.004	.001	.020
$f_Y(y)$	.910	.045	.032	.013	1.000

Definition 5.1.3 (Continuous joint density). Let X and Y be continuous random variables. The ordered pair (X, Y) is called a two-dimensional continuous random variable. A function f_{XY} such that

1. $f_{XY}(x, y) \geq 0 \qquad -\infty < x < \infty$
   $\qquad\qquad\qquad\quad -\infty < y < \infty$
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = 1$
3. $P[a \leq X \leq b \text{ and } c \leq Y \leq d] = \displaystyle\int_a^b \int_c^d f_{XY}(x, y)\, dy\, dx$

for a, b, c, d real is called the joint density for (X, Y).

Even though the joint density is defined for all real values x and y, we shall follow the convention of specifying its equation only over those regions for which it may be nonzero. Recall that in the case of a single continuous random variable, probabilities correspond to areas. In the case of a two-dimensional continuous random variable, probabilities correspond to *volumes*. These ideas are illustrated in Example 5.1.3.

Example 5.1.3. In a healthy individual age 20 to 29 years, the calcium level in the blood, X, is usually between 8.5 and 10.5 milligrams per deciliter (mg/dl) and the cholesterol level, Y, is usually between 120 and 240 mg/dl. Assume that for a healthy individual in this age group the random variable (X, Y) is uniformly distributed over the rectangle whose corners are (8.5, 120), (8.5, 240), (10.5, 120), (10.5, 240). That is, assume that the joint density for (X, Y) is

$$f_{XY}(x, y) = c \qquad 8.5 \leq x \leq 10.5$$
$$120 \leq y \leq 240$$

To be a density, c must be chosen so that

$$\int_{8.5}^{10.5}\int_{120}^{240} c\, dy\, dx = 1$$

That is, c must be chosen so that the volume of the rectangular solid shown in Fig. 5.1(a) is 1. To find c, we can use geometry or complete the indicated integration as shown below.

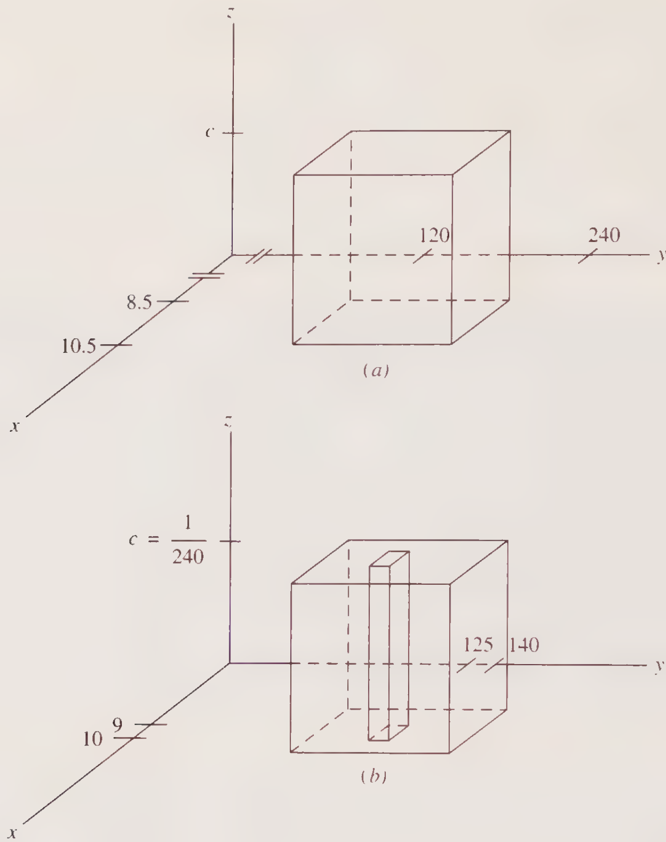


FIGURE 5.1
(*a*) Volume of the solid whose base is a rectangle with corners (8.5, 120), (8.5, 240), (10.5, 120), and (10.5, 240) and height c is 1; (*b*) $P[9 \leq X \leq 10 \text{ and } 125 \leq Y \leq 140]$ = volume of solid whose base is a rectangle with corners (9, 125), (9, 140), (10, 125), (10, 140) and height $c = 1/240$.

$$\int_{8.5}^{10.5} \int_{120}^{240} c \, dy \, dx = 1$$

$$c \int_{8.5}^{10.5} (240 - 120) \, dx = 1$$

$$120c(10.5 - 8.5) = 1$$

$$240c = 1$$

$$c = 1/240$$

Let us now use the joint density to find the probability that an individual's calcium level will lie between 9 and 10 mg/dl, whereas the cholesterol level is between 125 and 140 mg/dl. This probability corresponds to the volume of the solid shown in Fig. 5.1(*b*). This probability is

$$
\begin{aligned}
P[9 \le X \le 10 \text{ and } 125 \le Y \le 140] &= \int_9^{10} \int_{125}^{140} 1/240 \, dy \, dx \\
&= 1/240 \int_9^{10} (140 - 125) \, dx \\
&= 15/240
\end{aligned}
$$

To define "marginal" densities in the continuous case, we replace summation by integration. This yields the following definition.

Definition 5.1.4 (Continuous marginal densities). Let (X, Y) be a two-dimensional continuous random variable with joint density f_{XY}. The marginal density for X, denoted by f_X, is given by

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

The marginal density for Y, denoted by f_Y, is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) \, dx$$

We illustrate the idea of marginal densities in Examples 5.1.4 and 5.1.5.

Example 5.1.4. Let X denote an individual's blood calcium level and Y his or her blood cholesterol level. The joint density for (X, Y) is

$$f_{XY}(x, y) = 1/240 \qquad \begin{matrix} 8.5 \le x \le 10.5 \\ 120 \le y \le 240 \end{matrix}$$

The marginal densities for X and Y are

$$f_X(x) = \int_{120}^{240} 1/240 \, dy = 1/2 \qquad 8.5 \le x \le 10.5$$

$$f_Y(y) = \int_{8.5}^{10.5} 1/240 \, dx = 2/240 \qquad 120 \le y \le 240$$

To find the probability that a healthy individual has a cholesterol level between 150 and 200, we can use either the joint density or the marginal density for Y. That is,

$$P[150 \le Y \le 200] = \int_{8.5}^{10.5} \int_{150}^{200} 1/240 \, dy \, dx = 100/240$$

or

$$P[150 \le Y \le 200] = \int_{150}^{200} 2/240 \, dy = 100/240$$

Note that both X and Y are uniformly distributed.

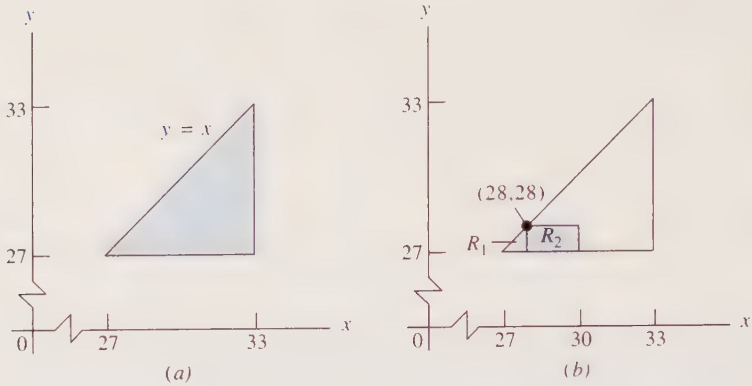


FIGURE 5.2
(a) The joint density $f(x, y) = c/x$ is defined over the triangular region bounded by $y = 27$, $y = x$, and $x = 33$.
(b)

$$P[X \le 30 \text{ and } Y \le 28] = \int\int_{R_1} c/x \, dy\, dx + \int\int_{R_2} c/x \, dy\, dx$$
$$= \int_{27}^{28}\int_{27}^{x} c/x \, dy\, dx + \int_{28}^{30}\int_{27}^{28} c/x \, dy\, dx$$

or

$$P[X \le 30 \text{ and } Y \le 28] = \int_{27}^{28}\int_{y}^{30} c/x \, dx\, dy.$$

Example 5.1.5. In studying the behavior of air support roofs, the random variables X, the inside barometric pressure (in inches of mercury), and Y, the outside pressure, are considered. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$
$$c = 1/(6 - 27 \ln 33/27) \doteq 1.72$$

The region in the plane over which this joint density is defined is shown in Fig. 5.2(a). The marginal densities for X and Y are given by

$$f_X(x) = \int_{27}^{x} c/x \, dy = (c/x)y \Big|_{27}^{x} = c(1 - 27/x) \qquad 27 \le x \le 33$$

$$f_Y(y) = \int_{y}^{33} c/x \, dx = c(\ln 33 - \ln y) \qquad 27 \le y \le 33$$

Let us find the probability that the inside pressure is at most 30 and the outside pressure is at most 28. That is, let us find $P[X \le 30 \text{ and } Y \le 28]$. The region over which the joint density is to be integrated is shown in Fig. 5.2(b). Integration can be done with respect to y and then x or vice versa. In the former case the problem must be split into two pieces, since the boundaries for y change at the point (28, 28). In the latter case integration can be accomplished more easily. The integrals required in the two cases are

Case I:

$$P[X \le 30 \text{ and } Y \le 28] = \int_{27}^{28}\int_{27}^{x} c/x\,dy\,dx + \int_{28}^{30}\int_{27}^{28} c/x\,dy\,dx$$

Case II:

$$P[X \le 30 \text{ and } Y \le 28] = \int_{27}^{28}\int_{y}^{30} c/x\,dx\,dy$$

Since case II requires less effort, we find $P[X \le 30 \text{ and } Y \le 28]$ as follows:

$$\begin{aligned}
P[X \le 30 \text{ and } Y \le 28] &= \int_{27}^{28}\int_{y}^{30} c/x\,dx\,dy \\
&= c\int_{27}^{28} [\ln 30 - \ln y]dy \\
&= c\left[y \ln 30\Big|_{27}^{28} - \int_{27}^{28} \ln y\,dy\right] \\
&= c\left[\ln 30 - (y \ln y - y)\Big|_{27}^{28}\right] \\
&= c[\ln 30 - 28 \ln 28 + 27 \ln 27 + 1] \\
&\doteq c(.09) = 1.72(.09) = .15
\end{aligned}$$

It is left as an exercise to show that the same result is obtained via case I. (See Exercise 6.)

Independence

There is one other point to be made in this section. Recall that two events are independent if knowledge of the fact that one has occurred gives us no clue as to the likelihood that the other will occur. Suppose that X and Y are discrete random variables such that knowledge of the value assumed by one gives us no clue as to the value assumed by the other. We would like to think of these random variables as being "independent" and would like a mathematical characterization of this property. The characterization is suggested by the following argument. Let X and Y be discrete. Let A_1 denote the event that $X = x$, and let A_2 denote the event that $Y = y$. If X and Y are independent in the intuitive sense, then A_1 and A_2 are independent events. By Definition 2.3.1

$$P[A_1 \cap A_2] = P[A_1]P[A_2]$$

Substituting, we see that

$$P[X = x \text{ and } Y = y] = P[X = x]P[Y = y]$$

or

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

It seems that, at least in the discrete case, independence implies that the *joint density can be expressed as the product of the marginal densities.* This idea provides the basis for the definition of the term "independent random variables" in both the discrete and continuous cases.

Definition 5.1.5 (Independent random variables). Let X and Y be random variables with joint density f_{XY} and marginal densities f_X and f_Y, respectively. X and Y are independent if and only if

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

for all x and y.

Example 5.1.6

(*a*) The random variables X, the number of defective welds, and Y, the number of improperly tightened bolts per car of Examples 5.1.1 and 5.1.2, are not independent. To verify this, note that from Table 5.2

$$f_{XY}(0, 0) = .84 \neq .9(.91) = .819 = f_X(0)f_Y(0)$$

(*b*) The random variables X, an individual's blood calcium level, and Y, his or her blood cholesterol level as described in Examples 5.1.3 and 5.1.4, are independent. To verify this, note that

$$f_{XY}(x, y) = 1/240 = 1/2 \cdot 2/240 = f_X(x)f_Y(y)$$

An important point should be made here. The assumption that (X, Y) is uniformly distributed leads to the conclusion that X and Y are independent. If this conclusion is *medically unsound*, then another more realistic density should be sought to describe the behavior of the two-dimensional random variable (X, Y).

(*c*) The random variables X and Y, the inside and outside pressure, respectively, on an air support roof of Example 5.1.5 are not independent. This is seen by noting that

$$f_{XY}(x, y) = c/x \neq c(1 - 27/x)c(\ln 33 - \ln y) = f_X(x)f_Y(y)$$

The assumption of nonindependence here is realistic from a physical point of view.

The exercises for Sec. 5.1 provide some practice in dealing with these theoretical ideas. You will see their relationship to data analysis in chapters to come.

5.2 EXPECTATION AND COVARIANCE

In this section we introduce the idea of *expectation* in the case of a two-dimensional random variable. We also study a specific expectation, called the *covariance*, that is useful in describing the behavior of one variable relative to another.

We begin by extending Definitions 3.3.1 and 4.2.1 to the two-dimensional case.

Definition 5.2.1 (Expected value). Let (X, Y) be a two-dimensional random variable with joint density f_{XY}. Let $H(X, Y)$ be a random variable. The expected value of $H(X, Y)$, denoted by $E[H(X, Y)]$ is given by

1. $E[H(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} H(x, y) f_{XY}(x, y)$

   provided $\sum_{\text{all } x} \sum_{\text{all } y} |H(x, y)| f_{XY}(x, y)$ exists for (X, Y) discrete;

2. $E[H(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f_{XY}(x, y) dy\, dx$

   provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |H(x, y)| f_{XY}(x, y) dy\, dx$ exists for (X, Y) continuous.

As in the case of one-dimensional random variables, some functions of X and Y are of more interest than others. In particular, if the joint density for (X, Y) is known, then the average value of X and of Y can be found easily. These are determined as follows:

Univariate Averages Found Via the Joint Density

$$E[X] = \sum_{\text{all } x} \sum_{\text{all } y} x f_{XY}(x, y) \qquad \text{for } (X, Y) \text{ discrete}$$

$$E[Y] = \sum_{\text{all } x} \sum_{\text{all } y} y f_{XY}(x, y)$$

$$E[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dx\, dy \qquad \text{for } (X, Y) \text{ continuous}$$

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

Examples 5.2.1 and 5.2.2 illustrate the use of this definition.

Example 5.2.1. The joint density for the random variable (X, Y) of Example 5.1.1 is given in Table 5.3. X denotes the number of defective welds and Y, the number of improperly tightened bolts produced per car by assembly line robots. Let us use Definition 5.2.1 to find $E[X]$, $E[Y]$, $E[X + Y]$, and $E[XY]$.

$$\begin{aligned} E[X] &= \sum_{x=0}^{2} \sum_{y=0}^{3} x f_{XY}(x, y) \\ &= 0(.840) + 0(.030) + 0(.020) + 0(.010) + 1(.060) + \cdots + 2(.001) \\ &= .12 \end{aligned}$$

TABLE 5.3

x/y	0	1	2	3	$f_X(x)$
0	.840	.030	.020	.010	.900
1	.060	.010	.008	.002	.080
2	.010	.005	.004	.001	.020
$f_Y(y)$	.910	.045	.032	.013	1.000

$$\begin{aligned} E[Y] &= \sum_{x=0}^{2} \sum_{y=0}^{3} y f_{XY}(x, y) \\ &= 0(.840) + 1(.030) + 2(.020) + 3(.010) + 0(.060) + \cdots + 3(.001) \\ &= .148 \end{aligned}$$

$$\begin{aligned} E[X + Y] &= \sum_{x=0}^{2} \sum_{y=0}^{3} (x + y) f_{XY}(x, y) \\ &= (0 + 0)(.840) + (0 + 1)(.030) + (0 + 2)(.020) + \cdots + (2 + 3)(.001) \\ &= .268 \end{aligned}$$

$$\begin{aligned} E[XY] &= \sum_{x=0}^{2} \sum_{y=0}^{3} xy f_{XY}(x, y) \\ &= (0 \cdot 0)(.840) + (0 \cdot 1)(.030) + (0 \cdot 2)(.020) + \cdots + (2 \cdot 3)(.001) \\ &= .064 \end{aligned}$$

There are two points to be made. First, both $E[X]$ and $E[Y]$ were found via the joint density and Definition 5.2.1. These expectations could have been found just as easily from the marginal densities and Definition 3.3.1. (See Exercise 18.) Second, note that $E[X + Y] = E[X] + E[Y]$. This result is consistent with the rules of expectation given in Theorem 3.3.1.

Example 5.2.2. The joint density for the random variable (X, Y), where X denotes the calcium level and Y denotes the cholesterol level in the blood of a healthy individual, is given by

$$f_{XY}(x, y) = 1/240 \qquad \begin{aligned} 8.5 \le x \le 10.5 \\ 120 \le y \le 240 \end{aligned}$$

For these variables,

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y)\, dy\, dx \\ &= \int_{8.5}^{10.5} \int_{120}^{240} x(1/240)\, dy\, dx \\ &= \int_{8.5}^{10.5} (1/2)x\, dx = x^2/4 \Big|_{8.5}^{10.5} = 9.5 \text{ mg/dl} \end{aligned}$$

$$
\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x, y)\, dy\, dx \\
&= \int_{8.5}^{10.5}\int_{120}^{240} y(1/240)\, dy\, dx \\
&= 1/240 \int_{8.5}^{10.5} y^2/2 \Big|_{120}^{240} dx \\
&= 1/240 \int_{8.5}^{10.5} 21{,}600\, dx = 180 \text{ mg/dl}
\end{aligned}
$$

$$
\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\, dy\, dx \\
&= \int_{8.5}^{10.5}\int_{120}^{240} xy(1/240)\, dy\, dx \\
&= 1/240 \int_{8.5}^{10.5} xy^2/2 \Big|_{120}^{240} dx \\
&= 1/240 \int_{8.5}^{10.5} 21{,}600x\, dx \\
&= (21{,}600/240)(x^2/2) \Big|_{8.5}^{10.5} = 1710
\end{aligned}
$$

Covariance

Occasionally the expected value of a function of X and Y is of interest in its own right. For instance, in Example 5.2.1, $E[X + Y]$ gives the theoretical average number of errors made by the robots overall. However, we shall be concerned primarily with those expectations that are needed to compute the covariance between X and Y. This term is defined as follows:

Definition 5.2.2 (Covariance). Let X and Y be random variables with means μ_X and μ_Y respectively. The covariance between X and Y, denoted by $\text{Cov}(X, Y)$ or σ_{XY} is given by

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

Note that if small values of X tend to be associated with small values of Y and large values of X with large values of Y, then $X - \mu_X$ and $Y - \mu_Y$ will usually have the same algebraic signs. This implies that $(X - \mu_X)(Y - \mu_Y)$ will be positive, yielding a positive covariance. If the reverse is true and small values of X tend to be associated with large values of Y and vice versa, then $X - \mu_X$ and $Y - \mu_Y$ will usually have opposite algebraic signs. This results in a negative value for $(X - \mu_X)(Y - \mu_Y)$, yielding a negative covariance. In this sense covariance is an indication of how X and Y vary relative to one another.

Covariance is seldom computed from Definition 5.2.2. Rather, we apply the following computational formula whose derivation is left as an exercise. (See Exercise 24.)

Theorem 5.2.1 (Computational formula for covariance)

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

We illustrate the use of Theorem 5.2.1 by finding the covariance for the random variables of Examples 5.2.1 and 5.2.2.

Example 5.2.3

(*a*) The covariance between X, the number of defective welds, and Y, the number of improperly tightened bolts of Example 5.2.1, is given by

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= .064 - (.12)(.148) = .046\end{aligned}$$

Since $\mathrm{Cov}(X, Y) > 0$, there is a tendency for large values of X to be associated with large values of Y and vice versa. That is, a car with an above average number of defective welds tends also to have an above average number of improperly tightened bolts and vice versa.

(*b*) The covariance between X, an individual's blood calcium level, and Y, his or her blood cholesterol level, has covariance given by

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\ &= 1710 - (9.5)(180) = 0\end{aligned}$$

A covariance of 0 implies that knowledge that X assumes a value above its mean gives us no indication as the value of Y relative to its mean.

The fact that the covariance between X and Y is 0 in Example 5.2.2 is not a coincidence. It is, of course, due to the fact that $E[XY] = E[X]E[Y]$. It can be shown that this property will hold whenever the random variables X and Y are independent, as they are in Example 5.2.2. This important result is formalized in the following theorem:

Theorem 5.2.2. Let (X, Y) be a two-dimensional random variable with joint density f_{XY}. If X and Y are independent then

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

Proof. We shall prove this theorem in the continuous case. The proof in the discrete case is similar. Assume that (X, Y) has joint density f_{XY} and that X and Y are independent. Let f_X and f_Y denote the marginal densities for X and Y, respectively. By Definition 5.2.1,

TABLE 5.4

x/y	-2	-1	1	2	$f_X(x)$
1	0	1/4	1/4	0	1/2
4	1/4	0	0	1/4	1/2
$f_Y(y)$	1/4	1/4	1/4	1/4	1

$$
\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\,dy\,dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dy\,dx \qquad (X \text{ and } Y \text{ are independent}) \\
&= \int_{-\infty}^{\infty} x f_X(x) \int_{-\infty}^{\infty} y f_Y(y)\,dy\,dx \\
&= \int_{\infty}^{\infty} x f_X(x) E[Y]\,dx \\
&= E[Y] \int_{-\infty}^{\infty} x f_X(x)\,dx = E[Y]E[X]
\end{aligned}
$$

An immediate consequence of this theorem is the result that we have already noted and observed relative to Example 5.2.2. In particular, *if X and Y are independent, then* $\mathrm{Cov}(X, Y) = 0$. Unfortunately, the converse of this statement is not true. That is, we *cannot* conclude that a zero covariance implies independence. The next example verifies this contention.

Example 5.2.4. The joint density for (X, Y) is given in Table 5.4, from which we see that $E[X] = 5/2$, $E[Y] = 0$, and $E[XY] = 0$, yielding a covariance of 0. It is also easy to see that X and Y are *not* independent. The value assumed by Y does have an effect on that assumed by X. In fact, $X = Y^2$. The value of Y completely determines the value of X!

Covariance gives us only a very rough idea of the relationship between X and Y. We are concerned only with its algebraic sign and not with its magnitude. However, covariance is used to define another measure of the relationship between X and Y which is easier to interpret. This measure, called the *correlation,* is discussed in the next section.

5.3 CORRELATION

Recall that the covariance between X and Y gives only a rough indication of any association that may exist between X and Y. No attempt is made to describe the type or strength of the association. Often it is of interest to know whether or not two random variables are *linearly* related. One measure used to determine this is the Pearson coefficient of correlation, ρ. In this section we define this theoretical measure of linearity; in Chap. 11 we shall discuss how to estimate its value from a data set.

Definition 5.3.1 (Pearson coefficient of correlation). Let X and Y be random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, respectively. The correlation, ρ_{XY}, between X and Y is given by

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{(\operatorname{Var} X)(\operatorname{Var} Y)}}$$

Since we already know how to calculate each of the terms appearing in the above definition, calculating ρ_{XY} (or ρ) from the joint density for (X, Y) is easy. The question is, "How do we interpret ρ once we know its numerical value?" To interpret ρ, we must know its range of possible values. The next theorem shows that, unlike the covariance which can assume any real value, the correlation coefficient is bounded.

Theorem 5.3.1. The correlation coefficient ρ_{XY} for any two random variables X and Y lies between -1 and 1 inclusive.

The proof of this theorem is found in Appendix C.

The next theorem indicates how ρ measures linearity. The point of the theorem is twofold. First, if there is a linear relationship between X and Y, then this fact is reflected in a correlation coefficient of 1 or -1. Second, if $\rho = 1$ or -1, then a linear relationship exists between X and Y. The formal statement of this result is given in Theorem 5.3.2.

Theorem 5.3.2. Let X and Y be random variables with correlation coefficient ρ_{XY}. Then $|\rho_{XY}| = 1$ if and only if $Y = \beta_0 + \beta_1 X$ for some real numbers β_0 and $\beta_1 \neq 0$.

See Appendix C for the proof of this theorem.

If $\rho = 1$, then we say that X and Y have *perfect positive* correlation. Perfect positive correlation implies that $Y = \beta_0 + \beta_1 X$, where $\beta_1 > 0$. This in turn implies that small values of X are associated with small values of Y, and large values of X with large values of Y. Perfect negative correlation implies that $Y = \beta_0 + \beta_1 X$, where $\beta_1 < 0$. Practically speaking, this means that small values of X are associated with large values of Y and vice versa. Unfortunately, random variables seldom assume the easily interpretable values of 1 or -1. However, values of ρ near 1 or -1 do occur and indicate a linear trend. That is, they indicate that, even though no single straight line passes through the points of positive probability, there is a straight line passing through the graph with the property that most of the probability is associated with points lying on or near this straight line. It is equally important to realize what Theorem 5.3.2 is not saying. If $\rho = 0$, we say that X and Y are uncorrelated, but we are *not* saying that they are unrelated. We are saying that if a relationship exists, then it is *not linear*. These ideas are illustrated in Fig. 5.3.

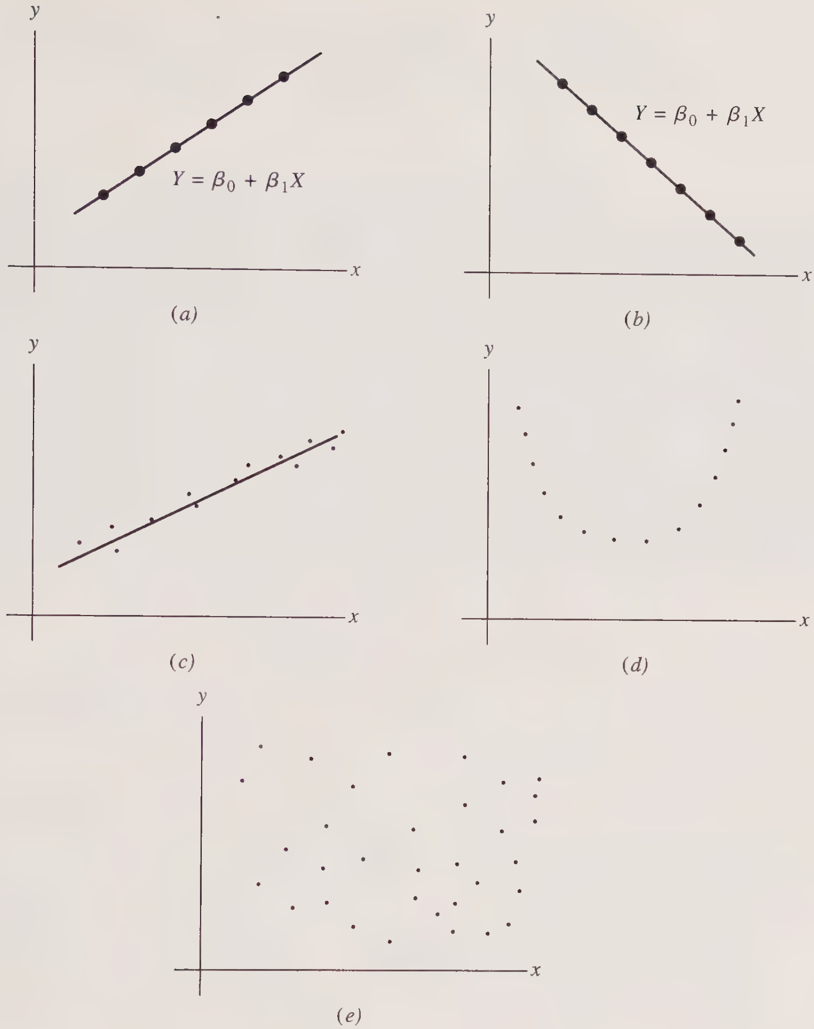


FIGURE 5.3
(*a*) Perfect positive correlation: $\rho = 1$, $\beta_1 > 0$, all points lie on a straight line with positive slope; (*b*) perfect negative correlation: $\rho = -1$, $\beta_1 < 0$, all points lie on a straight line with negative slope; (*c*) ρ near 1, points exhibit a linear trend; (*d*) uncorrelated: $\rho = 0$, points indicate a relationship between X and Y, but the relationship is not linear; (*e*) uncorrelated: $\rho = 0$, points are randomly scattered.

Example 5.3.1. To find the correlation between X, the number of defective welds, and Y, the number of improperly tightened bolts produced per car by assembly line robots, we use Table 5.3 to compute $E[X^2]$ and $E[Y^2]$. For these variables

$$E[X^2] = 0^2(.90) + 1^2(.08) + 2^2(.02) = .16$$
$$E[Y^2] = 0^2(.910) + 1^2(.045) + 2^2(.032) + 3^2(.013) = .29$$

In Example 5.2.1, we found that $E[X] = .12$ and $E[Y] = .148$. Therefore

$$\text{Var } X = E[X^2] - (E[X])^2 = .16 - (.12)^2 \doteq .146$$
$$\text{Var } Y = E[Y^2] - (E[Y])^2 = .29 - (.148)^2 \doteq .268$$

In Example 5.2.3 we found that Cov$(X, Y) = .046$. By Definition 5.3.1,

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X \text{ Var } Y}} = \frac{.046}{\sqrt{(.146)(.268)}} \doteq .23$$

Since this value does not appear to lie close to 1, we would not expect the observed values of X and Y to exhibit a strong linear trend.

Exercise 36 points out the relationship between correlation and independence.

5.4 CONDITIONAL DENSITIES AND REGRESSION

In this section we consider two topics that are closely related. These are *conditional densities* and *regression.* To see what is to be done, let us reconsider Example 5.1.5.

Example 5.4.1. In Example 5.1.5 we considered the random variable (X, Y) where X is the inside and Y the outside barometric pressure on an air support roof. Suppose we are interested in studying the inside pressure when the outside pressure is fixed at $y = 30$. There are three important points to understand:

1. The inside pressure will vary even though the outside pressure is constant. Therefore it makes sense to talk about "the random variable X given that $y = 30$." We shall denote this new random variable by $X|y = 30$.
2. Since $X|y = 30$ is a random variable in its own right, it has a probability distribution. Therefore it makes sense to ask, "What is the density for $X|y = 30$?" We shall call this density the "conditional density for X given that $y = 30$" and shall denote it by $f_{X|y=30}$.
3. Since the inside pressure varies even though the outside pressure is constant, it makes sense to ask, "What is the mean or average pressure on the inside of the roof when the outside pressure is 30?" That is, we can ask, "What is the mean value for the random variable $X|y = 30$?" This mean value is denoted by $E[X|y = 30]$ or $\mu_{X|y=30}$.

In general, the conditional density for X given $Y = y$, denoted by $f_{X|y}$, is a function that allows us to find the probability that X assumes specific values based on knowledge of the value assumed by the random variable Y. To see how to define $f_{X|y}$ let us assume that (X, Y) is discrete with joint density f_{XY} and marginal densities f_X and f_Y. Let A_1 denote the event that $X = x$ and A_2 denote the event that $Y = y$. From Definition 2.2.1,

$$P[A_1|A_2] = \frac{P[A_1 \cap A_2]}{P[A_2]}$$

Substituting, we see that

$$P[X = x|Y = y] = \frac{P[X = x \text{ and } Y = y]}{P[Y = y]} = \frac{f_{XY}(x, y)}{f_Y(y)}$$

In the discrete case the conditional density for X given $Y = y$ is the ratio of the joint density for (X, Y) to the marginal density for Y. This observation provides the motivation for the definition of the term "conditional density" in both the discrete and continuous cases. In the formal definition, note that the roles of X and Y can be reversed.

Definition 5.4.1 (Conditional density). Let (X, Y) be a two-dimensional random variable with joint density f_{XY} and marginal densities f_X and f_Y. Then

1. The conditional density for X given $Y = y$, denoted by $f_{X|y}$ is given by

$$f_{X|y} = \frac{f_{XY}(x, y)}{f_Y(y)} \qquad f_Y(y) > 0$$

2. The conditional density for Y given $X = x$, denoted by $f_{Y|x}$, is given by

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \qquad f_X(x) > 0$$

The use of this definition is illustrated in Example 5.4.2.

Example 5.4.2. The joint density for the random variable (X, Y), where X is the inside and Y is the outside pressure on an air support roof, is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$
$$c = 1/(6 - 27 \ln 33/27)$$

From Example 5.1.5 the marginal densities for X and Y are

$$f_X(x) = c(1 - 27/x) \qquad 27 \le x \le 33$$

and

$$f_Y(y) = c(\ln 33 - \ln y) \qquad 27 \le y \le 33$$

The conditional density for X given $Y = y$ is

$$\begin{aligned} f_{X|y}(x) &= \frac{f_{XY}(x, y)}{f_Y(y)} \\ &= \frac{c/x}{c(\ln 33 - \ln y)} = \frac{1}{x(\ln 33 - \ln y)} \qquad y \le x \le 33 \end{aligned}$$

To find the probability that the inside pressure exceeds 32 given that the outside pressure is 30, we let $y = 30$ in the above expression. We then integrate the conditional density over values of X that exceed 32. That is,

$$\begin{aligned} P[X > 32 | y = 30] &= \int_{32}^{33} \frac{1}{x(\ln 33 - \ln 30)}\, dx \\ &= \left. \frac{\ln x}{\ln 33 - \ln 30} \right|_{32}^{33} \\ &= \frac{\ln 33 - \ln 32}{\ln 33 - \ln 30} \doteq .32 \end{aligned}$$

To find the expected or mean value of X given $y = 30$ we apply Definition 4.2.1 to the random variable $X|\,y = 30$. That is,

$$
\begin{aligned}
E[X|\,y = 30] = \mu_{X|y=30} &= \int_{-\infty}^{\infty} x f_{X|\,y=30}\,dx \\
&= \int_{30}^{33} x \frac{1}{x(\ln 33 - \ln 30)}\,dx \\
&= \int_{30}^{33} \frac{1}{\ln 33 - \ln 30}\,dx \\
&= \frac{3}{\ln 33 - \ln 30} \doteq 31.48
\end{aligned}
$$

When the outside pressure on the roof is 30, the average value of the inside pressure is 31.48 inches of mercury.

Curves of Regression

In the previous example, note that we did not find the mean for X. We found the mean for X when $y = 30$. The mean value obtained depended on the value chosen for Y. In general, the mean of X given $Y = y$ or $\mu_{X|y}$ is a *function of* y. When this function is graphed, we obtain what is called the *curve of regression of X on Y*. This term is defined formally in Definition 5.4.2. Note that, once again, the roles of X and Y can be reversed.

Definition 5.4.2 (Curve of regression). Let (X, Y) be a two-dimensional random variable.

1. The graph of the mean value of X given $Y = y$, denoted by $\mu_{X|y}$, is called the curve of regression of X on Y.
2. The graph of the mean value of Y given $X = x$, denoted by $\mu_{Y|x}$, is called the curve of regression of Y on X.

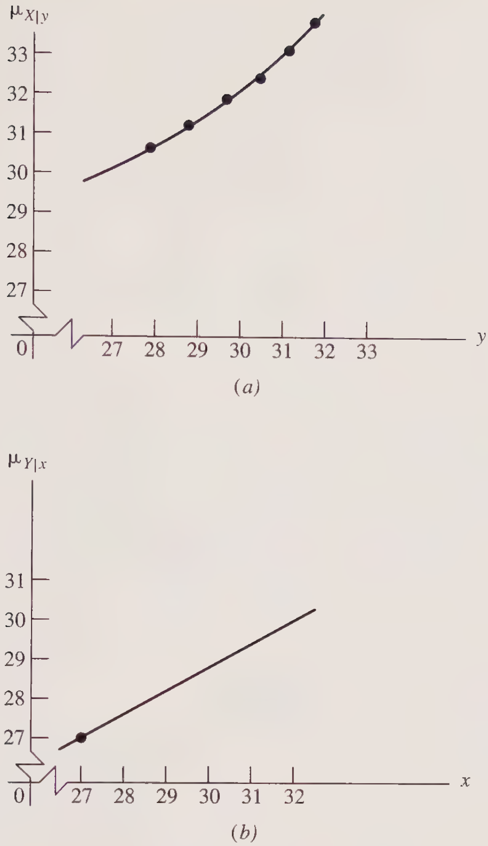
We illustrate the use of this definition by finding the curve of regression of X on Y and the curve of regression of Y on X for the random variable (X, Y) of Example 5.4.2.

Example 5.4.3. The conditional density for X given $Y = y$, where X is the inside and Y is the outside pressure on an air support roof, is given by

$$
f_{X|y}(x) = \frac{1}{x(\ln 33 - \ln y)} \qquad y \le x \le 33
$$

The equation for the curve of regression of X on Y is given by

$$
\begin{aligned}
\mu_{X|y} &= \int_{y}^{33} x \frac{1}{x(\ln 33 - \ln y)}\,dx \\
&= \int_{y}^{33} \frac{1}{\ln 33 - \ln y}\,dx \\
&= \frac{33 - y}{\ln 33 - \ln y}
\end{aligned}
$$



y	$\mu_{X\|y}$
27	29.90
28	30.43
29	30.95
30	31.48
31	31.98
32	32.50
32.5	32.74

FIGURE 5.4
(*a*) A nonlinear curve of regression: $\mu_{X|y} = (33 - y)/(\ln 33 - \ln y)$; (*b*) a linear curve of regression: $\mu_{Y|x} = (1/2)(x + 27)$.

Note that this equation is *nonlinear.* Its graph is not a straight line. A sketch of the graph is found by plotting $\mu_{X|y}$ for selected values of y. The graph is shown in Fig. 5.4(*a*). The conditional density for Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

$$= \frac{c/x}{c(1 - 27/x)}$$

$$= \frac{1}{x - 27} \qquad 27 \le y \le x$$

The equation for the curve of regression of Y on X is given by

$$\begin{aligned}
\mu_{Y|x} &= \int_{27}^{x} y \frac{1}{x-27}\,dy \\
&= \frac{y^2}{2(x-27)}\Big|_{27}^{x} \\
&= \frac{x^2-27^2}{2(x-27)} \\
&= (1/2)(x+27)
\end{aligned}$$

Note that this equation is *linear*. Its graph is the straight line shown in Fig. 5.4(b). These curves can be used now to find the mean of X for any specified value of Y or vice versa. For example, the average value of Y, the outside pressure, given that the inside pressure is 29 is

$$\mu_{Y|x=29} = (1/2)(x+27) = (1/2)(56) = 28 \text{ inches of mercury}$$

We have introduced only the basic ideas underlying the topic of regression. To find the theoretical regression curves, you must *know* the joint density for (X, Y). In practice, this density is seldom known with certainty. Thus, in practice, we are forced to approximate these theoretical curves from a data set—a set of observations on the random variable (X, Y). Methods for doing so are presented in Chaps. 11 and 12.

5.5 TRANSFORMATION OF VARIABLES

In Sec. 4.8 we considered the problem of transforming continuous variables in the univariate case. That is, given a continuous random variable X whose density is known, we saw how to find the density for the random variable Y, where Y is a function of X. Here we reconsider the problem in the bivariate case. To do so, we must first introduce the notation of Jacobians.

Suppose that we are working in the xy plane and that u and v are variables, each of which is a function of x and y. That is,

$$u = g_1(x, y) \qquad \text{and} \qquad v = g_2(x, y)$$

These two equations define a transformation T from some region in the xy plane into the uv plane, as pictured in Fig. 5.5(a). Assume that g_1 and g_2 have continuous partial derivatives with respect to x and y. The Jacobian of T is denoted by J_T and is given by the following determinant:

$$J_T = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\[2ex] \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}$$

Example 5.5.1 illustrates the idea.

Example 5.5.1. Consider the transformation T from the xy plane into the uv plane defined by

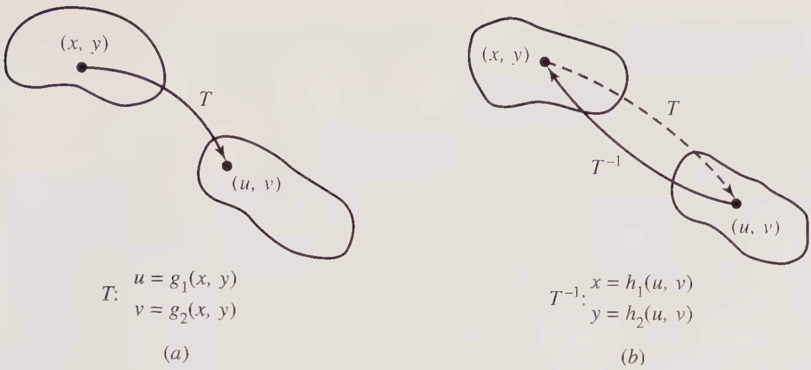


FIGURE 5.5
(*a*) T maps from the xy plane into the uv plane; (*b*) T^{-1} maps from the uv plane into the xy plane.

$$u = g_1(x, y) = (3y - x)/6$$
$$v = g_2(x, y) = x/3$$

The Jacobian of T is

$$J_T = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} -1/6 & 1/2 \\ 1/3 & 0 \end{vmatrix} = (-1/6)(0) - (1/2)(1/3) = -(1/6)$$

If a transformation T is one-to-one, then it is invertible. Assume that the inverse transformation, T^{-1}, is defined by the equations

$$x = h_1(u, v) \qquad \text{and} \qquad y = h_2(u, v)$$

and that h_1 and h_2 have continuous partial derivatives. [See Fig. 5.5(*b*).] The Jacobian of this inverse transformation is given by the determinant

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

This is the sort of Jacobian that will be useful to us in the statistical setting.

Assume that we have two continuous random variables X and Y whose joint density f_{XY} is known. Let U and V be random variables, each of which is a function of X and Y. We want to determine the form of f_{UV}, the joint density for (U, V), based on knowledge of the form of f_{XY}. The method for doing so parallels Theorem 4.8.1 and is given in Theorem 5.5.1.

Theorem 5.5.1. Let (X, Y) be continuous with joint density f_{XY}. Let

$$U = g_1(X, Y) \qquad \text{and} \qquad V = g_2(X, Y)$$

where g_1 and g_2 define a one-to-one transformation. Let the inverse transformation be defined by

$$X = h_1(U, V) \qquad \text{and} \qquad Y = h_2(U, V)$$

where h_1 and h_2 have continuous first partial derivatives. Then the joint density for (U, V) is given by

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J|$$

where $J \neq 0$ is the Jacobian of the inverse transformation. That is,

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

It is easy to see that Theorem 4.8.1 is a special case of this theorem with f_X corresponding to f_{XY}, $g^{-1}(y)$ playing the role of the inverse transformation, and $|dg^{-1}(y)/dy|$ being equivalent to the absolute value of the Jacobian of the inverse transformation.

Example 5.5.2. Assume that X and Y are independent uniformly distributed random variables over $(0, 2)$ and $(0, 3)$, respectively. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/6 \qquad \begin{array}{l} 0 < x < 2 \\ 0 < y < 3 \end{array}$$

Let $U = X - Y$ and $V = X + Y$. What is the joint density for (U, V)? To apply Theorem 5.5.1, we first note that the transformation

$$T: \begin{cases} U = X - Y \\ V = X + Y \end{cases}$$

is a linear transformation from the xy plane into the uv plane. A result from advanced calculus states that a linear transformation from two-dimensional space into two-dimensional space is one-to-one whenever the determinant of its matrix of coefficients is not zero. Here the determinant is

$$\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = (1)(1) - (1)(-1) = 2$$

so T is invertible. The inverse transformation is found by solving the above system of equations for X and Y. Here T^{-1} is given by

$$T^{-1}: \begin{cases} X = (U + V)/2 \\ Y = (V - U)/2 \end{cases}$$

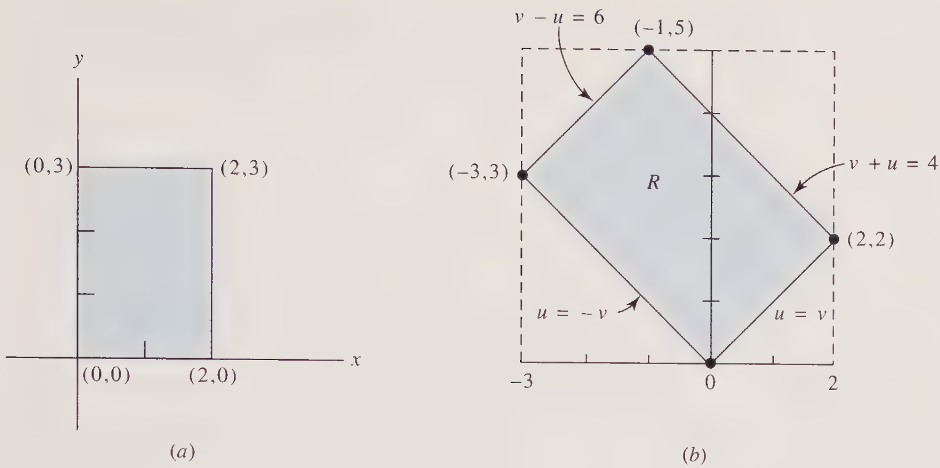


FIGURE 5.6
(a) (X, Y) lies in the rectangle with corners (0, 0), (2, 0), (0, 3), and (2, 3); (b) (U, V) lies in the region R.

The Jacobian of T^{-1} is

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{vmatrix} = (1/2)(1/2) - (1/2)(-1/2) = 1/2$$

By Theorem 5.5.1,

$$\begin{aligned} f_{UV}(u, v) &= f_{XY}((v + u)/2, (v - u)/2)|J| \\ &= (1/6)(1/2) = 1/12 \end{aligned}$$

To find the set of values for which $f_{UV} > 0$, we note that since $0 < x < 2$ and $0 < y < 3$, (X, Y) lies in the rectangle shown in Fig. 5.6(a). It is easy to see that $U = X - Y$ must lie between -3 and 2 and that $V = X + Y$ must lie between 0 and 5. Furthermore, U and V must satisfy the inequalities

$$\begin{aligned} 0 < (v + u)/2 < 2 \\ 0 < (v - u)/2 < 3 \end{aligned}$$

or

$$\begin{aligned} 0 < v + u < 4 \\ 0 < v - u < 6 \end{aligned}$$

Solving these inequalities simultaneously yields the region R shown in Fig. 5.6(b). Thus the density for (U, V) is given by

$$f_{UV}(u, v) = 1/12 \qquad (u, v) \in R$$

We leave it to you to verify that f_{UV} is, in fact, a valid density.

Other transformation theorems can be derived from Theorem 5.5.1. Some of these are given in Exercises 48, 50, and 51. For a more detailed discussion of this topic, please see [49].

CHAPTER SUMMARY

In this chapter we considered random variables of more than one dimension. Emphasis was on random variables of two dimensions. The joint density was defined by extending the notion of a density for a single variable in a logical way. This function was used to calculate probabilities associated with two-dimensional random variables (X, Y). We saw how to obtain the marginal densities for both X and Y from the joint density. These marginal densities are the usual densities for X or Y when considered alone. The correlation coefficient ρ was introduced as a measure of linearity between X and Y. The notion of independence between X and Y was defined formally, and its relationship to ρ was investigated. We saw how to define the conditional densities for X given Y and Y given X from knowledge of the joint density for (X, Y) and the marginal densities for X and Y. The conditional densities were used to find the equations for the curves of regression of Y on X and X on Y. These regression curves are the graphs of the mean value of Y as a function of X or vice versa. We saw that these curves may be linear or nonlinear.

We introduced and defined important terms that you should know. These are:

Two-dimensional discrete random variable
Two-dimensional continuous random variable
Discrete joint density
Discrete marginal density
Independent random variables
Covariance
Perfect positive correlation
Uncorrelated
Curve of regression
n-dimensional discrete random variable
n-dimensional continuous random variable
Bivariate normal distribution
Continuous joint density
Continuous marginal density
Expected value of $H(X, Y)$
Correlation coefficient
Perfect negative correlation
Conditional density

EXERCISES

Section 5.1

1. Use Table 5.2 to find each of these probabilities:
- (*a*) The probability that exactly two defective welds and one improperly tightened bolt will be produced by the robots.
- (*b*) The probability that at least one defective weld and at least one improperly tightened bolt will be produced.
- (*c*) The probability that at most one defective weld will be produced.
- (*d*) The probability that at least two improperly tightened bolts will be produced.

TABLE 5.5

x/y	0	1	2	3	4
0	0	0	0	0	1/35
1	0	0	0	12/35	0
2	0	0	18/35	0	0
3	0	4/35	0	0	0

2. In conducting an experiment in the laboratory, temperature gauges are to be used at four junction points in the equipment setup. These four gauges are randomly selected from a bin containing seven such gauges. Unknown to the scientist, three of the seven gauges give improper temperature readings. Let X denote the number of defective gauges selected and Y the number of nondefective gauges selected. The joint density for (X, Y) is given in Table 5.5.
   (*a*) The values given in Table 5.5 can be derived by realizing that the random variable X is hypergeometric. Use the results of Sec. 3.7 to verify the values given in Table 5.5.
   (*b*) Find the marginal densities for both X and Y. What type of random variable is Y?
   (*c*) Intuitively speaking, are X and Y independent? Justify your answer mathematically.

3. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/n^2 \qquad x = 1, 2, 3, \ldots, n$$
$$y = 1, 2, 3, \ldots, n$$

   (*a*) Verify that $f_{XY}(x, y)$ satisfies the conditions necessary to be a density.
   (*b*) Find the marginal densities for X and Y.
   (*c*) Are X and Y independent?

4. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 2/n(n + 1) \qquad 1 \leq y \leq x \leq n \qquad n \text{ a positive integer}$$

   (*a*) Verify that $f_{XY}(x, y)$ satisfies the conditions necessary to be a density. *Hint:* The sum of the first n integers is given by $n(n + 1)/2$.
   (*b*) Find the marginal densities for X and Y. *Hint:* Draw a picture of the region over which (X, Y) is defined.
   (*c*) Are X and Y independent?
   (*d*) Assume that $n = 5$. Use the joint density to find $P[X \leq 3 \text{ and } Y \leq 2]$. Find $P[X \leq 3]$ and $P[Y \leq 2]$. *Hint:* Draw a picture of the region over which (X, Y) is defined.

5. The two most common types of errors made by programmers are syntax errors and errors in logic. For a simple language such as BASIC the number of such errors is usually small. Let X denote the number of syntax errors and Y the number of errors in logic made on the first run of a BASIC program. Assume that the joint density for (X, Y) is as shown in Table 5.6.

TABLE 5.6

x/y	0	1	2	3
0	.400	.100	.020	.005
1	.300	.040	.010	.004
2	.040	.010	.009	.003
3	.009	.008	.007	.003
4	.008	.007	.005	.002
5	.005	.002	.002	.001

(*a*) Find the probability that a randomly selected program will have neither of these types of errors.
(*b*) Find the probability that a randomly selected program will contain at least one syntax error and at most one error in logic.
(c) Find the marginal densities for X and Y.
(*d*) Find the probability that a randomly selected program contains at least two syntax errors.
(*e*) Find the probability that a randomly selected program contains one or two errors in logic.
(*f*) Are X and Y independent?

6. Consider Example 5.1.5. Verify that $P[X \leq 30 \text{ and } Y \leq 28] \doteq .15$ by integrating the joint density first with respect to y, then with respect to x.

7. (*a*) Use the joint density of Example 5.1.5 to find the probability that the inside pressure on the roof will be greater than 30, and the outside pressure is less than 32.
(*b*) Use the marginal density for X to find $P[X \leq 28]$.
(c) Use the marginal density for Y to find $P[Y > 30]$.

8. Let X denote the temperature (°C) and let Y denote the time in minutes that it takes for the diesel engine on an automobile to get ready to start. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c(4x + 2y + 1) \qquad 0 \leq x \leq 40$$
$$0 \leq y \leq 2$$

(*a*) Find the value of c that makes this a density.
(*b*) Find the probability that on a randomly selected day the air temperature will exceed 20° C and it will take at least 1 minute for the car to be ready to start.
(*c*) Find the marginal densities for X and Y.
(*d*) Find the probability that on a randomly selected day it will take at least one minute for the car to be ready to start.
(*e*) Find the probability that on a randomly selected day the air temperature will exceed 20° C.
(*f*) Are X and Y independent? Explain on a mathematical basis.

9. An engineer is studying early morning traffic patterns at a particular intersection. The observation period begins at 5:30 a.m. Let X denote the time of arrival of the first vehicle from the north-south direction; let Y denote the first arrival time from the east-west direction. Time is measured in fractions of an hour after 5:30 a.m. Assume that the density for (X, Y) is given by

$$f_{XY}(x, y) = 1/x \qquad 0 < y < x < 1$$

(*a*) Verify that this is a joint density for a two-dimensional random variable.
(*b*) Find $P[X \leq .5 \text{ and } Y \leq .25]$.
(*c*) Find $P[X > .5 \text{ or } Y > .25]$.
(*d*) Find $P[X \geq .5 \text{ and } Y \geq .5]$.
(*e*) Find the marginal densities for X and Y.
(*f*) Find $P[X \leq .5]$.
(*g*) Find $P[Y \leq .25]$.
(*h*) Are X and Y independent? Explain.

10. The joint density for (X, Y) is given by

$$f_{XY}(x, y) = x^3y^3/16 \qquad 0 \leq x \leq 2, 0 \leq y \leq 2$$

(*a*) Find the marginal densities for X and Y.
(*b*) Are X and Y independent?
(*c*) Find $P[X \leq 1]$.
(*d*) If it is known that $y = 1$, what is $P[X \leq 1]$? (Do not use any computation to answer this question!)

11. Economic conditions cause fluctuations in the prices of raw commodities as well as in finished products. Let X denote the price paid for a barrel of crude oil by the initial carrier, and let Y denote the price paid by the refinery purchasing the product from the carrier. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = c \qquad 20 < x < y < 40$$

(*a*) Find the value of c that makes this a joint density for a two-dimensional random variable.
(*b*) Find the probability that the carrier will pay at least \$25 per barrel and the refinery will pay at most \$30 per barrel for the oil.
(*c*) Find the probability that the price paid by the refinery exceeds that of the carrier by at least \$10 per barrel.
(*d*) Find the marginal densities for X and Y.
(*e*) Find the probability that the price paid by the carrier is at least \$25.
(*f*) Find the probability that the price paid by the refinery is at most \$30.
(*g*) Are X and Y independent? Explain.

12. (*n-dimensional discrete random variables.*) Random variables of dimension $n > 2$ can be defined and studied by extending the definitions presented in the two-dimensional case in a logical way. For example, an n-tuple $(X_1, X_2, X_3, \ldots, X_n)$ in which each of the random variables $X_1, X_2, X_3, \ldots, X_n$ is a discrete random variable is called an n-dimensional discrete random variable. The density for such a random variable is given by

$$f(x_1, x_2, x_3, \ldots, x_n) = P[X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n]$$

This problem entails the use of a three-dimensional random variable.

Items coming off an assembly line are classed as being either nondefective, defective but salvageable, or defective and nonsalvageable. The probabilities of observing items in each of these categories are .9, .08, and .02, respectively. The probabilities do not change from trial to trial. Twenty items are randomly selected and classified. Let X_1 denote the number of nondefective items obtained, X_2 the number of defective but salvageable items obtained, and X_3 the number of defective and nonsalvageable items obtained.

(*a*) Find $P[X_1 = 15, X_2 = 3, X_3 = 2]$. *Hint:* Use the formula for the number of permutations of indistinguishable objects, page 16, Chap. 1, to count the number of ways to get this sort of split in a sequence of 20 trials.

(*b*) Find the general formula for the density for (X_1, X_2, X_3).

13. (*n-dimensional continuous random variables.*) An n-tuple $(X_1, X_2, X_3, \ldots, X_n)$, where each of the random variables $X_1, X_2, \ldots, X_n$ is continuous, is called an n-dimensional continuous random variable. The density for an n-dimensional continuous random variable is defined by extending Definition 5.1.3 in a natural way. State the three properties that identify a function as a density for $(X_1, X_2, X_3, \ldots, X_n)$.

14. Let $f(x_1, x_2, x_3) = c(x_1 \cdot x_2 \cdot x_3)$ for $0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1$. Find the value of c that makes this a density for the three-dimensional random variable (X_1, X_2, X_3).

Section 5.2

15. Four temperature gauges are randomly selected from a bin containing three defective and four nondefective gauges. Let X denote the number of defective gauges selected and Y the number of nondefective gauges selected. (See Exercise 2.) The joint density for (X, Y) is given in Table 5.5.

(*a*) From the physical description of the problem, should $\mathrm{Cov}(X, Y)$ be positive or negative?

(*b*) Find $E[X]$, $E[Y]$, $E[XY]$, and $\mathrm{Cov}(X, Y)$.

16. Let X denote the number of syntax errors and Y the number of errors in logic made on the first run of a BASIC program. (See Exercise 5.) The joint density for (X, Y) is given in Table 5.6.

(*a*) X and Y are not independent. Does this give any indication of the value of the covariance?

(*b*) Find $E[X]$, $E[Y]$, $E[XY]$, and $\mathrm{Cov}(X, Y)$. Give a rough physical interpretation of the covariance.

(*c*) Find $E[X + Y]$. What is the practical interpretation of this expectation?

17. Consider the random variable (X, Y) of Exercise 3. Without doing any additional computation, find $\mathrm{Cov}(X, Y)$.

18. Use the marginal densities given in Table 5.3 to compute $E[X]$ and $E[Y]$. Compare your results to those obtained in Example 5.2.1.

19. The joint density for (X, Y), where X is the inside and Y is the outside barometric pressure on an air support roof (see Example 5.1.5), is given by

$$f_{XY}(x, y) = c/x \qquad 27 \le y \le x \le 33$$

$$c = 1/(6 - 27 \ln 33/27) \doteq 1.72$$

(a) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.
(b) Find $E[X - Y]$. What is the practical physical interpretation of this expectation?

20. The joint density for (X, Y), where X is the temperature and Y is the time that it takes for a diesel engine on an automobile to get ready to start (see Exercise 8), is given by

$$f_{XY}(x, y) = (1/6640)(4x + 2y + 1) \qquad 0 \le x \le 40$$
$$0 \le y \le 2$$

(a) From a physical standpoint, do you think $\text{Cov}(X, Y)$ should be positive or negative?
(b) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.

21. The joint density for (X, Y), where X is the arrival time of the first vehicle from the north-south direction and Y is the arrival time of the first vehicle from the east-west direction at an intersection (see Exercise 9), is given by

$$f_{XY}(x, y) = 1/x \qquad 0 < y < x < 1$$

Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.

22. Find the covariance between the random variables X and Y of Exercise 10.

23. Let X denote the price paid for a barrel of crude oil by the initial carrier, and let Y denote the price paid by the refinery purchasing the oil. (See Exercise 11.) The joint density for (X, Y) is given by

$$f_{XY}(x, y) = 1/200 \qquad 20 < x < y < 40$$

(a) From a physical standpoint, should $\text{Cov}(X, Y)$ be positive or negative?
(b) Find $E[X]$, $E[Y]$, $E[XY]$, and $\text{Cov}(X, Y)$.
(c) Find $E[Y - X]$. Interpret this expectation in a practical sense.

24. Show that $\text{Cov}(XY) = E[XY] - E[X]E[Y]$. *Hint:* By definition, $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$. Expand this product, and apply the rules for expectation (Theorem 3.3.1). Remember that $\mu_X = E[X]$ and $\mu_Y = E[Y]$.

25. Prove that $\text{Var}(X + Y) = \text{Var}\, X + \text{Var}\, Y + 2\,\text{Cov}(X, Y)$. *Hint:* $\text{Var}(X + Y) = E[(X + Y)^2] - (E[X + Y])^2$. Square these terms, and apply the rules for expectation. (Theorem 3.3.1.)

26. Use the result of Exercise 25 to show that if X and Y are independent, then $\text{Var}(X + Y) = \text{Var}\, X + \text{Var}\, Y$. This proves the third rule for variance. (Theorem 3.3.3.)

27. Show that if $X = Y$, then $\text{Cov}(X, Y) = \text{Var}\, X = \text{Var}\, Y$.

28. Let the joint density for (X, Y) be given by

$$f(x, y) = \frac{1}{2(e-1)}\left[\frac{1}{x} + \frac{1}{y}\right] \qquad 1 \le x \le e \qquad 1 \le y \le e$$

(a) Show that $\int_1^e \int_1^e f(x, y)\, dy\, dx = 1$.
(b) Find $E[X]$ and $E[Y]$.
(c) Find $E[XY]$.
(d) Are X and Y independent? Explain, based on your answers to parts (b) and (c) and Theorem 5.2.2.

Section 5.3

29. The joint density for (X, Y), where X denotes the number of defective and Y the number of nondefective temperature gauges selected from a bin containing three defective and four nondefective gauges, is given in Table 5.5. (See Exercise 2.)

(*a*) From the physical interpretation of the problem, should ρ_{XY} be positive or negative? Should ρ_{XY} be $+1$ or -1? Explain.

(*b*) Find $E[X^2]$ and $E[Y^2]$. Use the information from Exercise 15 to find ρ_{XY}.

In Exercises 30 to 34, find $E[X^2]$, $E[Y^2]$, Var X, Var Y, and ρ_{XY} for the random variables in the exercises referenced. In each case decide whether or not you would expect the graph of Y versus X to exhibit a strong linear trend.

30. Exercise 16.
31. Exercise 19.
32. Exercise 20.
33. Exercise 21.
34. Exercise 23.
35. Assume that $Y = \beta_0 + \beta_1 X$, $\beta_1 \neq 0$.

(*a*) Show that $\text{Cov}(X, Y) = \beta_1 \text{ Var } X$. *Hint:* $\text{Cov}(X, Y) = E[X(\beta_0 + \beta_1 X)] - E[X]E[\beta_0 + \beta_1 X]$. Use the rules for expectation.

(*b*) Show that Var $Y = \beta_1^2$ Var X. *Hint:* Use the rules for variance. (Theorem 3.3.3.)

(*c*) Find ρ_{XY}.

(*d*) Argue that $\rho_{XY} = 1$ if β_1, the slope of the line $Y = \beta_0 + \beta_1 X$, is positive and that $\rho_{XY} = -1$ if the slope of this line is negative.

36. Prove that if X and Y are independent, then $\rho_{XY} = 0$. Can we conclude that if X and Y are uncorrelated, then they are independent? Explain.

37. Without doing any additional computation, find ρ_{XY} for the random variables of Exercise 3.

38. What is the correlation between the random variables X and Y of Exercise 10?

Section 5.4

39. Consider Example 5.4.3.

(*a*) What is the expected value of X when $y = 31$?

(*b*) What is the expected value of Y when $x = 30$?

40. Consider Example 5.1.4.

(*a*) Find $f_{X|y}$. Note that $f_{X|y} = f_X$. From a physical standpoint, can you explain why these densities are the same?

(*b*) Find $f_{Y|x}$. Is $f_{Y|x} = f_Y$?

(*c*) Find the curve of regression of X on Y and the curve of regression of Y on X. Are these curves linear?

41. Consider the random variable (X, Y) of Exercise 4.

(*a*) Find the curve of regression of X on Y. Is the regression linear?

(*b*) Assume that $n = 10$ and find the mean value of X when $y = 4$.

(*c*) Find the curve of regression of Y on X. Is the regression linear?

(*d*) Assume that $n = 10$ and find the mean value of Y when $x = 4$.

42. Consider the random variable (X, Y) of Exercise 9.
(*a*) Find the curve of regression of X on Y. Is the regression linear?
(*b*) Find the mean value of X when $y = .5$.
(*c*) Find the curve of regression of Y on X. Is the regression linear?
(*d*) Find the mean value of Y when $x = .75$.

43. Consider Exercise 11.
(*a*) Find the curve of regression of X on Y. Is the regression linear?
(*b*) Find the mean price paid by the carrier for a barrel of crude oil given that the refinery price is \$30 per barrel.
(*c*) Find the curve of regression of Y on X. Is the regression linear?
(*d*) Find the mean price paid by the refinery for a barrel of crude oil given that the carrier paid \$35 per barrel.

44. Note that if $|\rho| = 1$, then $Y = \beta_0 + \beta_1 X$. For fixed values of X, $Y|x = \beta_0 + \beta_1 x$. Argue that $\mu_{Y|x}$ is a linear function of x. That is, argue that if X and Y are perfectly correlated, then the curve of regression of Y on X is linear. Is the converse true? Explain.

Section 5.5

45. Consider the linear transformation T defined by

$$T\colon u = 2x + y$$
$$v = x + 3y$$

(*a*) Is this transformation invertible? If so, find the defining equations for T^{-1}.
(*b*) Find the Jacobian for T^{-1}.

46. Consider the linear transformation T defined by

$$T\colon u = 3x + 2y$$
$$v = x - y$$

(*a*) Is this transformation invertible? If so, find the defining equations for T^{-1}.
(*b*) Find the Jacobian for T^{-1}.

47. Assume that X and Y are independent and uniformly distributed over $(0, 1)$ and $(0,2)$, respectively. Find the joint density for (U, V), where U and V are as defined in Exercise 45.

48. *(Distribution of one function of two continuous random variables.)* Let X and Y be continuous random variables with joint density f_{XY}. Let $U = X + Y$. Prove that f_U, the density for $X + Y$, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v)\, dv$$

Hint: Define a transformation T by

$$u = g_1(x, y) = x + y$$
$$v = g_2(x, y) = y$$

Follow the procedure given in Theorem 5.5.1 to obtain the joint density for (U, V). Integrate the joint density to obtain the marginal density for U.

49. Let X and Y be independent standard normal random variables. Let $U = X + Y$. Use Exercise 48 to prove that U follows a normal distribution with mean 0 and variance 2. *Hint:* In integrating over v, complete the square in the exponent and remember that a normal density integrated over the real line is equal to 1.

50. Let X and Y be continuous random variables with joint density f_{XY}. Let $U = XY$. Prove that f_U, the density for XY, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u/v, v)\,|1/v|\,dv$$

Hint: Let $u = g_1(x, y) = xy$ and $v = y$, and apply Theorem 5.5.1.

51. Let X and Y be continuous random variables with joint density f_{XY}. Let $U = X/Y$. Prove that f_U, the density for X/Y, is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v)\,|v|\,dv$$

Hint: Let $u = g_1(x, y) = x/y$ and $v = y$, and apply Theorem 5.5.1.

52. Let X and Y be independent exponentially distributed random variables with parameters β_1 and β_2, respectively.

(*a*) Find the joint density for (X, Y).

(*b*) Let $U = X + Y$, and verify that

$$f_U(u) = \int_0^u f_{XY}(u - v, v)\,dv$$

Hint: Remember that $0 < x < \infty$ and that $x = u - v$.

(*c*) Assume that $\beta_1 = 3$ and $\beta_2 = 1$. Show that

$$f_U(u) = e^{-u/3} - e^{-u/2} \qquad 0 < u < \infty$$

53. Let X and Y be independent uniformly distributed random variables over the intervals $(0, 2)$ and $(0, 3)$, respectively.

(*a*) Let $U = XY$ and find f_U.

(*b*) Let $U = X/Y$ and find f_U.

REVIEW EXERCISES

54. An electronic device is designed to switch house lights on and off at random times after it has been activated. Assume that the device is designed in such a way that it will be switched on and off exactly once in a 1-hour period. Let Y denote the time at which the lights are turned on and X the time at which they are turned off. Assume that the joint density for (X, Y) is given by

$$f_{XY}(x, y) = 8xy \qquad 0 < y < x < 1$$

(*a*) Verify that f_{XY} satisfies the conditions necessary to be a density.

(*b*) Find $E[XY]$.

(*c*) Find the probability that the lights will be switched on within 1/2 hour after being activated and then switched off again within 15 minutes.

TABLE 5.7

x/y	1	2	3	4
0	.059	.100	.050	.001
1	.093	.120	.082	.003
2	.065	.102	.100	.010
3	.050	.075	.070	.020

(*d*) Find the marginal density for X. Find $E[X]$ and $E[X^2]$.
(*e*) Find the marginal density for Y. Find $E[Y]$ and $E[Y^2]$.
(*f*) Are X and Y independent?
(*g*) Find the conditional distribution of X given Y.
(*h*) Find the probability that the lights will be switched off within 45 minutes of the system being activated given that they were switched on 10 minutes after the system was activated.
(*i*) Find the curve of regression of X on Y. Is the regression linear?
(*j*) Find the expected time that the lights will be turned off given that they were turned on 10 minutes after the system was activated.
(*k*) Based on the physical description of the problem, would you expect ρ to be positive, negative, or 0? Explain. Verify by computing ρ.

55. Verify that

$$f_{XY}(x, y) = xye^{-x}e^{-y} \qquad x > 0 \qquad y > 0$$

satisfies the conditions necessary to be a density for a continuous random variable (X, Y). Find the marginal densities for X and Y. Are X and Y independent? Find ρ_{XY}.

56. Let X denote the number of "do loops" in a Fortran program and Y the number of runs needed for a novice to debug the program. Assume that the joint density for (X, Y) is given in Table 5.7.
(*a*) Find the probability that a randomly selected program contains at most one "do loop" and requires at least two runs to debug the program.
(*b*) Find $E[XY]$.
(*c*) Find the marginal densities for X and Y. Use these to find the mean and variance for both X and Y.
(*d*) Find the probability that a randomly selected program requires at least two runs to debug given that it contains exactly one "do loop."
(*e*) Find Cov(X, Y). Find the correlation between X and Y. Based on the observed value of ρ, can you claim that X and Y are not independent? Explain.

57. Vehicles arrive at a highway toll booth at random instances from both the south and north. Assume that they arrive at average rates of five and three per 5-minute period, respectively. Let X denote the number arriving from the south during a 5-minute period, and let Y denote the number arriving from the north during this same time. Assume that X and Y are independent.
(*a*) Find the joint density for (X, Y).

(*b*) Find the probability that a total of four vehicles arrives during a five-minute time period.
(*c*) Find the correlation between X and Y.
(*d*) Find the conditional density for X given $Y = y$.

58. *(Bivariate normal distribution.)* A random variable (X, Y) is said to have a bivariate normal distribution if its joint density is given by

$$f_{XY}(x, y) = \frac{\exp\left\{\dfrac{1}{-2(1-\rho^2)}\left[\left(\dfrac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\dfrac{x-\mu_X}{\sigma_X}\right)\left(\dfrac{y-\mu_Y}{\sigma_Y}\right) + \left(\dfrac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

where x and y can assume any real value. The parameters μ_X, μ_Y, σ_X, σ_Y denote the respective means and standard deviations for X and Y. The parameter ρ is the correlation coefficient. The name of this distribution comes from the fact that the marginal densities for X and Y are both normal. Show that in the case of a bivariate normal distribution, if $\rho = 0$, then X and Y are independent.