CHAPTER
6
TWO RANDOM VARIABLES

6-1 BIVARIATE DISTRIBUTIONS

We are given two random variables **x** and **y**, defined as in Sec. 4-1, and we wish to determine their joint statistics, that is, the probability that the point (**x**, **y**) is in a specified region[1] D in the xy plane. The distribution functions $F_x(x)$ and $F_y(y)$ of the given random variables determine their separate (marginal) statistics but not their joint statistics. In particular, the probability of the event

$$\{\mathbf{x} \le x\} \cap \{\mathbf{y} \le y\} = \{\mathbf{x} \le x, \mathbf{y} \le y\}$$

cannot be expressed in terms of $F_x(x)$ and $F_y(y)$. Here, we show that the joint statistics of the random variables **x** and **y** are completely determined if the probability of this event is known for every x and y.

Joint Distribution and Density

The joint (bivariate) distribution $F_{xy}(x, y)$ or, simply, $F(x, y)$ of two random variables **x** and **y** is the probability of the event

$$\{\mathbf{x} \le x, \mathbf{y} \le y\} = \{(\mathbf{x}, \mathbf{y}) \in D_1\}$$

where x and y are two arbitrary real numbers and D_1 is the quadrant shown in Fig. 6-1a:

$$F(x, y) = P\{\mathbf{x} \le x, \mathbf{y} \le y\} \tag{6-1}$$

[1]The region D is arbitrary subject only to the mild condition that it can be expressed as a countable union or intersection of rectangles.

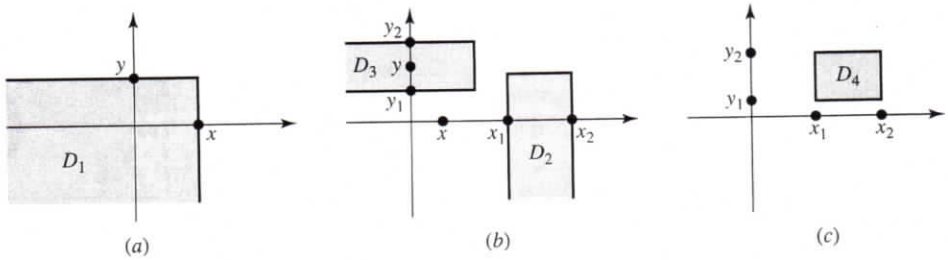

FIGURE 6-1

PROPERTIES

1. The function $F(x, y)$ is such that

$$F(-\infty, y) = 0, \quad F(x, -\infty) = 0, \quad F(\infty, \infty) = 1$$

Proof. As we know, $P\{\mathbf{x} = -\infty\} = P\{\mathbf{y} = -\infty\} = 0$. And since

$$\{\mathbf{x} = -\infty, \mathbf{y} \le y\} \subset \{\mathbf{x} = -\infty\} \qquad \{\mathbf{x} \le x, \mathbf{y} = -\infty\} \subset \{\mathbf{y} = -\infty\}$$

the first two equations follow. The last is a consequence of the identities

$$\{\mathbf{x} \le -\infty, \mathbf{y} \le -\infty\} = S \qquad P(S) = 1$$

2. The event $\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y\}$ consists of all points $(\mathbf{x}, \mathbf{y})$ in the vertical half-strip D_2 and the event $\{\mathbf{x} \le x, y_1 < \mathbf{y} \le y_2\}$ consists of all points $(\mathbf{x}, \mathbf{y})$ in the horizontal half-strip D_3 of Fig. 6-1b. We maintain that

$$\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y\} = F(x_2, y) - F(x_1, y) \tag{6-2}$$

$$\{\mathbf{x} \le x, y_1 < \mathbf{y} \le y_2\} = F(x, y_2) - F(x, y_1) \tag{6-3}$$

Proof. Clearly, for $x_2 > x_1$

$$\{\mathbf{x} \le x_2, \mathbf{y} \le y\} = \{\mathbf{x} \le x_1, \mathbf{y} \le y\} \cup \{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y\}$$

The last two events are mutually exclusive; hence [see (2-10)]

$$P\{\mathbf{x} \le x_2, \mathbf{y} \le y\} = P\{\mathbf{x} \le x_1, \mathbf{y} \le y\} + P\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y\}$$

and (6-2) results. The proof of (6-3) is similar.

3.

$$P\{x_1 < \mathbf{x} \le x_2, y_1 < \mathbf{y} \le y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \tag{6-4}$$

This is the probability that $(\mathbf{x}, \mathbf{y})$ is in the rectangle D_4 of Fig. 6-1c.

Proof. It follows from (6-2) and (6-3) because

$$\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y_2\} = \{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y_1\} \cup \{x_1 < \mathbf{x} \le x_2, y_1 < \mathbf{y} \le y_2\}$$

and the last two events are mutually exclusive.

JOINT DENSITY. The joint density of **x** and **y** is by definition the function

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{6-5}$$

From this and property 1 it follows that

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha, \beta)\, d\alpha\, d\beta \tag{6-6}$$

JOINT STATISTICS. We shall now show that the probability that the point $(\mathbf{x}, \mathbf{y})$ is in a region D of the xy plane equals the integral of $f(x, y)$ in D. In other words,

$$P\{(\mathbf{x}, \mathbf{y}) \in D\} = \int_D \int f(x, y)\, dx\, dy \tag{6-7}$$

where $\{(\mathbf{x}, \mathbf{y}) \in D\}$ is the event consisting of all outcomes ζ such that the point $[\mathbf{x}(\zeta), \mathbf{y}(\zeta)]$ is in D.

Proof. As we know, the ratio

$$\frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x\, \Delta y}$$

tends to $\partial F(x, y)/\partial x \partial y$ as $\Delta x \to 0$ and $\Delta y \to 0$. Hence [see (6-4) and (6-5)]

$$P\{x < \mathbf{x} \le x + \Delta x, y < \mathbf{y} \le y + \Delta y\} \simeq f(x, y)\, \Delta x\, \Delta y \tag{6-8}$$

We have thus shown that the probability that $(\mathbf{x}, \mathbf{y})$ is in a differential rectangle equals $f(x, y)$ times the area $\Delta x\, \Delta y$ of the rectangle. This proves (6-7) because the region D can be written as the limit of the union of such rectangles.

MARGINAL STATISTICS. In the study of several random variables, the statistics of each are called marginal. Thus $F_x(x)$ is the *marginal distribution* and $f_x(x)$ the *marginal density* of **x**. Here, we express the marginal statistics of **x** and **y** in terms of their joint statistics $F(x, y)$ and $f(x, y)$.

We maintain that

$$F_x(x) = F(x, \infty) \qquad F_y(y) = F(\infty, y) \tag{6-9}$$

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad f_y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{6-10}$$

Proof. Clearly, $\{\mathbf{x} \le \infty\} = \{\mathbf{y} \le \infty\} = S$; hence

$$\{\mathbf{x} \le x\} = \{\mathbf{x} \le x, \mathbf{y} \le \infty\} \qquad \{\mathbf{y} \le y\} = \{\mathbf{x} \le \infty, \mathbf{y} \le y\}$$

The probabilistics of these two sides yield (6-9).

Differentiating (6-6), we obtain

$$\frac{\partial F(x, y)}{\partial x} = \int_{-\infty}^{y} f(x, \beta)\, d\beta \qquad \frac{\partial F(x, y)}{\partial y} = \int_{-\infty}^{x} f(\alpha, y)\, d\alpha \tag{6-11}$$

Setting $y = \infty$ in the first and $x = \infty$ in the second equation, we obtain (6-10) because [see (6-9)]

$$f_x(x) = \frac{\partial F(x, \infty)}{\partial x} \qquad f_y(x) = \frac{\partial F(\infty, y)}{\partial y}$$

EXISTENCE THEOREM. From properties 1 and 3 it follows that

$$F(-\infty, y) = 0 \qquad F(x, -\infty) = 0 \qquad F(\infty, \infty) = 1 \tag{6-12}$$

and

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0 \tag{6-13}$$

for every $x_1 < x_2$, $y_1 < y_2$. Hence [see (6-6) and (6-8)]

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1 \qquad f(x, y) \geq 0 \tag{6-14}$$

Conversely, given $F(x, y)$ or $f(x, y)$ as before, we can find two random variables **x** and **y**, defined in some space S, with distribution $F(x, y)$ or density $f(x, y)$. This can be done by extending the existence theorem of Sec. 4-3 to joint statistics.

Probability Masses

The probability that the point (x, y) is in a region D of the plane can be interpreted as the probability mass in this region. Thus the mass in the entire plane equals 1. The mass in the half-plane $\mathbf{x} \leq x$ to the left of the line L_x of Fig. 6-2 equals $F_x(x)$. The mass in the half-plane $\mathbf{y} \leq y$ below the line L_y equals $F_y(y)$. The mass in the doubly-shaded quadrant $\{\mathbf{x} \leq x, \mathbf{y} \leq y\}$ equals $F(x, y)$.

Finally, the mass in the clear quadrant $(\mathbf{x} > x, \mathbf{y} > y)$ equals

$$P\{\mathbf{x} > x, \mathbf{y} > y\} = 1 - F_x(x) - F_y(y) + F(x, y) \tag{6-15}$$

The probability mass in a region D equals the integral [see (6-7)]

$$\int_D \int f(x, y)\, dx\, dy$$

If, therefore, $f(x, y)$ is a bounded function, it can be interpreted as surface mass density.

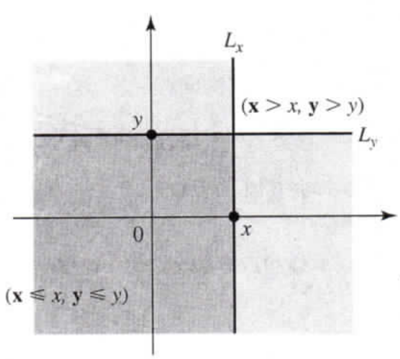

FIGURE 6-2

6-2 ONE FUNCTION OF TWO RANDOM VARIABLES

Given two random variables **x** and **y** and a function $g(x, y)$, we form a new random variable **z** as

$$\mathbf{z} = g(\mathbf{x}, \mathbf{y}) \tag{6-36}$$

Given the joint p.d.f. $f_{xy}(x, y)$, how does one obtain $f_z(z)$, the p.d.f. of **z**? Problems of this type are of interest from a practical standpoint. For example, a received signal in a communication scene usually consists of the desired signal buried in noise, and this formulation in that case reduces to $\mathbf{z} = \mathbf{x} + \mathbf{y}$. It is important to know the statistics of the incoming signal for proper receiver design. In this context, we shall analyze problems of the type shown in Fig. 6-6. Referring to (6-36), to start with,

$$\begin{aligned} F_z(z) &= P\{\mathbf{z}(\xi) \le z\} = P\{g(\mathbf{x}, \mathbf{y}) \le z\} = P\{(\mathbf{x}, \mathbf{y}) \in D_z\} \\ &= \iint_{x, y \in D_z} f_{xy}(x, y)\, dx\, dy \end{aligned} \tag{6-37}$$

where D_z in the xy plane represents the region where the inequality $g(x, y) \le z$ is satisfied (Fig. 6-7).

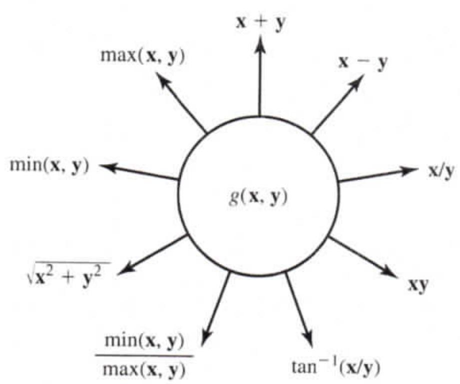

FIGURE 6-6

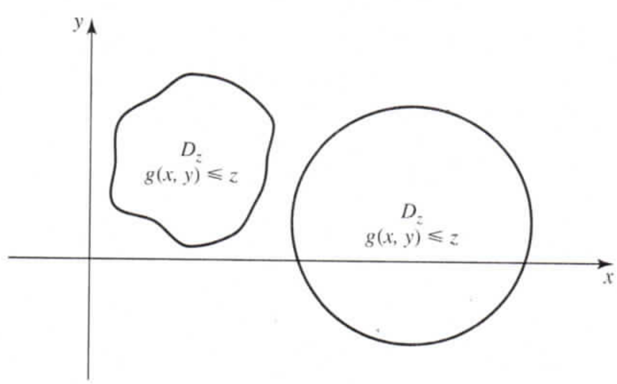

FIGURE 6-7

Note that D_z need not be simply connected. From (6-37), to determine $F_z(z)$ it is enough to find the region D_z for every z, and then evaluate the integral there.

We shall illustrate this method to determine the statistics of various functions of $\mathbf{x}$ and $\mathbf{y}$.

EXAMPLE 6-6

$\mathbf{z} = \mathbf{x} + \mathbf{y}$

▶ Let $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Determine the p.d.f. $f_z(z)$.

From (6-37),

$$F_z(z) = P\{\mathbf{x} + \mathbf{y} \le z\} = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_{xy}(x, y)\, dx\, dy \tag{6-38}$$

since the region D_z of the xy plane where $x + y \le z$ is the shaded area in Fig. 6-8 to the left of the line $x + y \le z$. Integrating over the horizontal strip along the x axis first (inner integral) followed by sliding that strip along the y axis from $-\infty$ to $+\infty$ (outer integral) we cover the entire shaded area.

We can find $f_z(z)$ by differentiating $F_z(z)$ directly. In this context it is useful to recall the differentiation rule due to Leibnitz. Suppose

$$F_z(z) = \int_{a(z)}^{b(z)} f(x, z)\, dx \tag{6-39}$$

Then

$$f_z(z) = \frac{dF_z(z)}{dz} = \frac{db(z)}{dz} f(b(z), z) - \frac{da(z)}{dz} f(a(z), z) + \int_{a(z)}^{b(z)} \frac{\partial f(x, z)}{\partial z}\, dx \tag{6-40}$$

Using (6-40) in (6-38) we get

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{xy}(x, y)\, dx \right) dy \\ &= \int_{-\infty}^{\infty} \left(1 \cdot f_{xy}(z - y, y) - 0 + \int_{-\infty}^{z-y} \frac{\partial f_{xy}(x, y)}{\partial z} \right) dy \\ &= \int_{-\infty}^{\infty} f_{xy}(z - y, y)\, dy \end{aligned} \tag{6-41}$$

Alternatively, the integration in (6-38) can be carried out first along the y axis followed by the x axis as in Fig. 6-9 as well (see problem set).

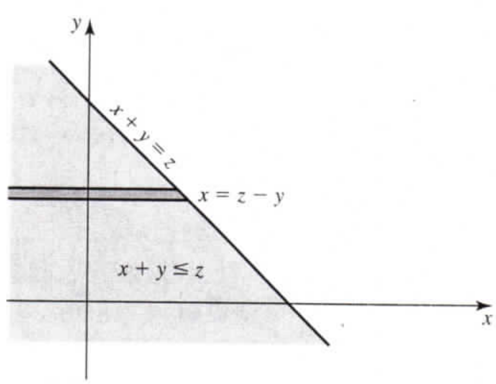

FIGURE 6-8

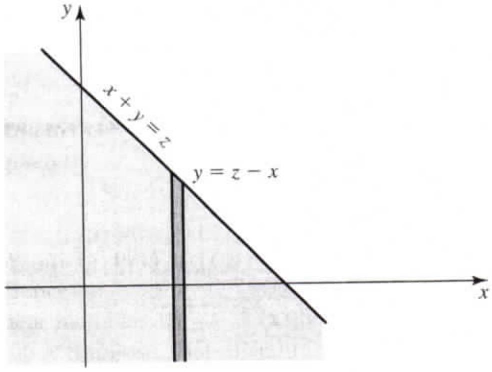

FIGURE 6-9

If **x** and **y** are independent, then

$$f_{xy}(x, y) = f_x(x)f_y(y) \tag{6-42}$$

and inserting (6-42) into (6-41) we get

$$f_z(z) = \int_{x=-\infty}^{\infty} f_x(z-y)f_y(y)\,dy = \int_{x=-\infty}^{\infty} f_x(x)f_y(z-x)\,dx \tag{6-43}$$

This integral is the convolution of the functions $f_x(z)$ and $f_y(z)$ expressed two different ways. We thus reach the following conclusion: If two random variables are *independent*, then the density of their sum equals the convolution of their densities.

As a special case, suppose that $f_x(x) = 0$ for $x < 0$ and $f_y(y) = 0$ for $y < 0$, then we can make use of Fig. 6-10 to determine the new limits for D_z.

In that case

$$F_z(z) = \int_{y=0}^{z}\int_{x=0}^{z-y} f_{xy}(x, y)\,dx\,dy$$

or

$$f_z(z) = \int_{y=0}^{z}\left(\frac{\partial}{\partial z}\int_{x=0}^{z-y} f_{xy}(x, y)\,dx\right)dy$$

$$= \begin{cases} \int_0^z f_{xy}(z-y, y)\,dy & z > 0 \\ 0 & z \le 0 \end{cases} \tag{6-44}$$

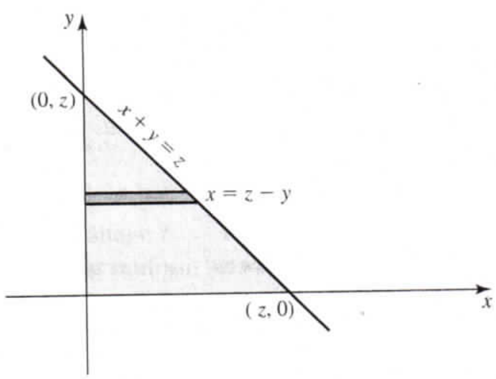

FIGURE 6-10

On the other hand, by considering vertical strips first in Fig. 6-10, we get

$$F_z(z) = \int_{x=0}^{z} \int_{y=0}^{z-x} f_{xy}(x, y)\, dy\, dx$$

or

$$f_z(z) = \int_{x=0}^{z} f_{xy}(x, z-x)\, dx$$

$$= \begin{cases} \int_0^z f_x(x) f_y(z-x)\, dx & z > 0 \\ 0 & z \le 0 \end{cases} \tag{6-45}$$

if **x** and **y** are independent random variables. ◀

EXAMPLE 6-7 ▶ Suppose **x** and **y** are independent exponential random variables with common parameter λ. Then

$$f_x(x) = \lambda e^{-\lambda x} U(x) \qquad f_y(y) = \lambda e^{-\lambda y} U(y) \tag{6-46}$$

and we can make use of (6-45) to obtain the p.d.f. of $\mathbf{z} = \mathbf{x} + \mathbf{y}$.

$$f_z(z) = \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)}\, dx = \lambda^2 e^{-\lambda z} \int_0^z dx$$

$$= z\lambda^2 e^{-\lambda z} U(z) \tag{6-47}$$

As Example 6-8 shows, care should be taken while using the convolution formula for random variables with finite range. ◀

EXAMPLE 6-8 ▶ **x** and **y** are independent uniform random variables in the common interval (0, 1). Determine $f_z(z)$, where $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Clearly,

$$\mathbf{z} = \mathbf{x} + \mathbf{y} \Rightarrow 0 < z < 2$$

and as Fig. 6-11 shows there are two cases for which the shaded areas are quite different in shape, and they should be considered separately.

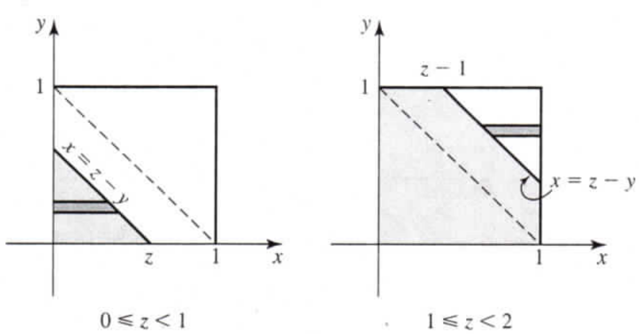

FIGURE 6-11

For $0 \le z < 1$,

$$F_z(z) = \int_{y=0}^{z}\int_{x=0}^{z-y} 1\,dx\,dy = \int_{y=0}^{z}(z-y)\,dy = \frac{z^2}{2} \qquad 0 < z < 1 \tag{6-48}$$

For $1 \le z < 2$, notice that it is easy to deal with the unshaded region. In that case,

$$F_z(z) = 1 - P\{\mathbf{z} > z\} = 1 - \int_{y=z-1}^{1}\int_{x=z-y}^{1} 1\,dx\,dy$$

$$= 1 - \int_{y=z-1}^{1}(1 - z + y)\,dy = 1 - \frac{(2-z)^2}{2} \qquad 1 \le z < 2 \tag{6-49}$$

Thus

$$f_z(z) = \frac{dF_z(z)}{dz} = \begin{cases} z & 0 \le z < 1 \\ 2 - z & 1 \le z < 2 \end{cases} \tag{6-50}$$

By direct convolution of $f_x(x)$ and $f_y(y)$, we obtain the same result as above. In fact, for $0 \le z < 1$ (Fig. 6-12a)

$$f_z(z) = \int f_x(z-x) f_y(x)\,dx = \int_0^z 1\,dx = z \tag{6-51}$$

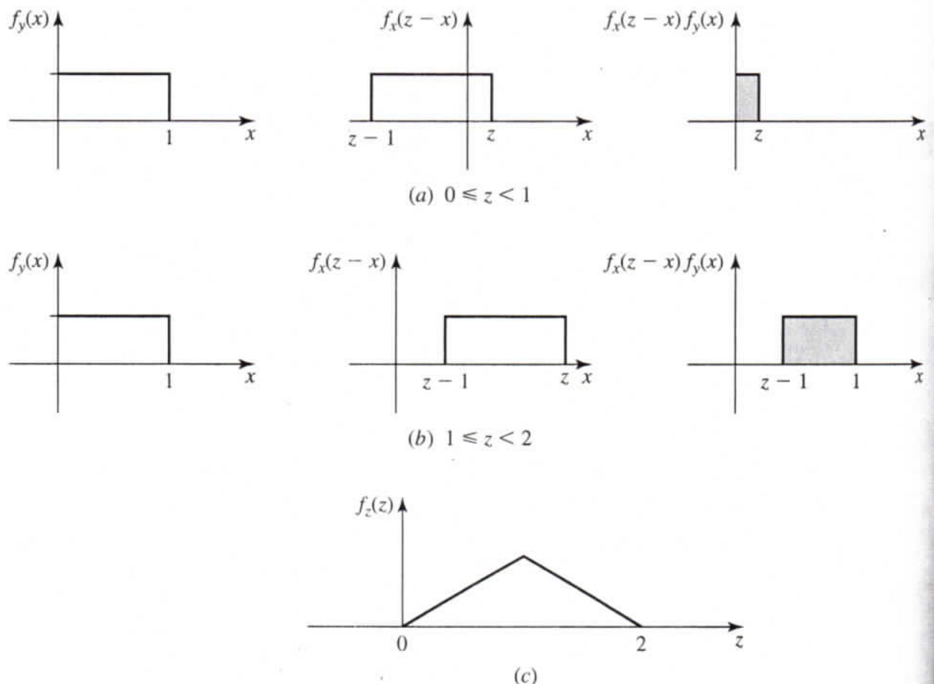

FIGURE 6-12

and for $1 \le z < 2$ (Fig. 6-12b)

$$f_z(z) = \int_{z-1}^{1} 1\,dx = 2 - z \tag{6-52}$$

Fig. 6-12c shows $f_z(z)$, which agrees with the convolution of two rectangular waveforms as well. ◀

EXAMPLE 6-9

$\mathbf{z} = \mathbf{x} - \mathbf{y}$

▶ Let $\mathbf{z} = \mathbf{x} - \mathbf{y}$. Determine $f_z(z)$.

From (6-37) and Fig. 6-13

$$F_z(z) = P\{\mathbf{x} - \mathbf{y} \le z\} = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z+y} f_{xy}(x, y)\,dx\,dy$$

and hence

$$f_z(z) = \frac{dF_z(z)}{dz} = \int_{-\infty}^{\infty} f_{xy}(z + y, y)\,dy \tag{6-53}$$

If $\mathbf{x}$ and $\mathbf{y}$ are independent, then this formula reduces to

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z + y) f_y(y)\,dy = f_x(-z) \otimes f_y(y) \tag{6-54}$$

which represents the convolution of $f_x(-z)$ with $f_y(z)$.

As a special case, suppose

$$f_x(x) = 0 \qquad x < 0, \qquad f_y(y) = 0 \qquad y < 0$$

In this case, $\mathbf{z}$ can be negative as well as positive, and that gives rise to two situations that should be analyzed separately, since the regions of integration for $z \ge 0$ and $z < 0$ are quite different.

For $z \ge 0$, from Fig. 6-14a

$$F_z(z) = \int_{y=0}^{\infty} \int_{x=0}^{z+y} f_{xy}(x, y)\,dx\,dy$$

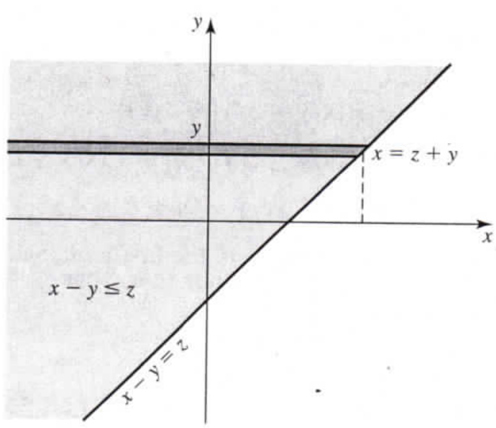

FIGURE 6-13

Differentiating with respect to w, we obtain

$$f_W(w) = \begin{cases} \lambda\mu/(\lambda+\mu w)^2 & w \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.56}$$

Quiz 4.6

(A) *Two computers use modems and a telephone line to transfer e-mail and Internet news every hour. At the start of a data call, the modems at each end of the line negotiate a speed that depends on the line quality. When the negotiated speed is low, the computers reduce the amount of news that they transfer. The number of bits transmitted L and the speed B in bits per second have the joint PMF*

$P_{L,B}(l,b)$	$b = 14,400$	$b = 21,600$	$b = 28,800$
$l = 518,400$	0.2	0.1	0.05
$l = 2,592,000$	0.05	0.1	0.2
$l = 7,776,000$	0	0.1	0.2

(4.57)

Let T denote the number of seconds needed for the transfer. Express T as a function of L and B. What is the PMF of T?

(B) *Find the CDF and the PDF of $W = XY$ when random variables X and Y have joint PDF*

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.58}$$

4.7 Expected Values

There are many situations in which we are interested only in the expected value of a derived random variable $W = g(X, Y)$, not the entire probability model. In these situations, we can obtain the expected value directly from $P_{X,Y}(x, y)$ or $f_{X,Y}(x, y)$ without taking the trouble to compute $P_W(w)$ or $f_W(w)$. Corresponding to Theorems 2.10 and 3.4, we have:

Theorem 4.12 *For random variables X and Y, the expected value of $W = g(X, Y)$ is*

$$\textit{Discrete:} \quad E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) P_{X,Y}(x,y),$$

$$\textit{Continuous:} \quad E[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy.$$

Example 4.11 In Example 4.8, compute $E[D]$ directly from $P_{L,T}(l,t)$.

Applying Theorem 4.12 to the discrete random variable D, we obtain

$$E[D] = \sum_{l=1}^{3} \sum_{t=40,60} lt P_{L,T}(l,t) \tag{4.59}$$
$$= (1)(40)(0.15) + (1)60(0.1) + (2)(40)(0.3) + (2)(60)(0.2) \tag{4.60}$$
$$+ (3)(40)(0.15) + (3)(60)(0.1) = 96 \text{ sec}, \tag{4.61}$$

which is the same result obtained in Example 4.8 after calculating $P_D(d)$.

Theorem 4.12 is surprisingly powerful. For example, it lets us calculate easily the expected value of a sum.

Theorem 4.13

$$E[g_1(X,Y) + \cdots + g_n(X,Y)] = E[g_1(X,Y)] + \cdots + E[g_n(X,Y)].$$

Proof Let $g(X,Y) = g_1(X,Y) + \cdots + g_n(X,Y)$. For discrete random variables X, Y, Theorem 4.12 states

$$E[g(X,Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} (g_1(x,y) + \cdots + g_n(x,y)) P_{X,Y}(x,y). \tag{4.62}$$

We can break the double summation into n double summations:

$$E[g(X,Y)] = \sum_{x \in S_X} \sum_{y \in S_Y} g_1(x,y) P_{X,Y}(x,y) + \cdots + \sum_{x \in S_X} \sum_{y \in S_Y} g_n(x,y) P_{X,Y}(x,y). \tag{4.63}$$

By Theorem 4.12, the ith double summation on the right side is $E[g_i(X,Y)]$, thus

$$E[g(X,Y)] = E[g_1(X,Y)] + \cdots + E[g_n(X,Y)]. \tag{4.64}$$

For continuous random variables, Theorem 4.12 says

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x,y) + \cdots + g_n(x,y)) f_{X,Y}(x,y)\, dx\, dy. \tag{4.65}$$

To complete the proof, we express this integral as the sum of n integrals and recognize that each of the new integrals is an expected value, $E[g_i(X,Y)]$.

In words, Theorem 4.13 says that the expected value of a sum equals the sum of the expected values. We will have many occasions to apply this theorem. The following theorem describes the expected sum of two random variables, a special case of Theorem 4.13.

Theorem 4.14 *For any two random variables X and Y,*

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

An important consequence of this theorem is that we can find the expected sum of two random variables from the separate probability models: $P_X(x)$ and $P_Y(y)$ or $f_X(x)$ and $f_Y(y)$. We do not need a complete probability model embodied in $P_{X,Y}(x,y)$ or $f_{X,Y}(x,y)$.

By contrast, the variance of $X + Y$ depends on the entire joint PMF or joint CDF:

Theorem 4.15 *The variance of the sum of two random variables is*

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2E\left[(X-\mu_X)(Y-\mu_Y)\right].$$

Proof Since $E[X+Y] = \mu_X + \mu_Y$,

$$\text{Var}[X+Y] = E\left[(X+Y-(\mu_X+\mu_Y))^2\right] \tag{4.66}$$

$$= E\left[((X-\mu_X)+(Y-\mu_Y))^2\right] \tag{4.67}$$

$$= E\left[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right]. \tag{4.68}$$

We observe that each of the three terms in the preceding expected values is a function of X and Y. Therefore, Theorem 4.13 implies

$$\text{Var}[X+Y] = E\left[(X-\mu_X)^2\right] + 2E\left[(X-\mu_X)(Y-\mu_Y)\right] + E\left[(Y-\mu_Y)^2\right]. \tag{4.69}$$

The first and last terms are, respectively, $\text{Var}[X]$ and $\text{Var}[Y]$.

The expression $E[(X-\mu_X)(Y-\mu_Y)]$ in the final term of Theorem 4.15 reveals important properties of the relationship of X and Y. This quantity appears over and over in practical applications, and it has its own name, *covariance*.

Definition 4.4 ***Covariance***
*The **covariance** of two random variables X and Y is*

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

Sometimes, the notation σ_{XY} is used to denote the covariance of X and Y. The *correlation* of two random variables, denoted $r_{X,Y}$, is a close relative of the covariance.

Definition 4.5 ***Correlation***
*The **correlation** of X and Y is $r_{X,Y} = E[XY]$*

The following theorem contains useful relationships among three expected values: the covariance of X and Y, the correlation of X and Y, and the variance of $X + Y$.

Theorem 4.16

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

(b) $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X,Y]$.

(c) If $X = Y$, $\text{Cov}[X,Y] = \text{Var}[X] = \text{Var}[Y]$ *and* $r_{X,Y} = E[X^2] = E[Y^2]$.

Theorem 4.18 *If X and Y are random variables such that $Y = aX + b$,*

$$\rho_{X,Y} = \begin{cases} -1 & a < 0, \\ 0 & a = 0, \\ 1 & a > 0. \end{cases}$$

The proof is left as an exercise for the reader (Problem 4.7.7). Some examples of positive, negative, and zero correlation coefficients include:

- X is the height of a student. Y is the weight of the same student. $0 < \rho_{X,Y} < 1$.
- X is the distance of a cellular phone from the nearest base station. Y is the power of the received signal at the cellular phone. $-1 < \rho_{X,Y} < 0$.
- X is the temperature of a resistor measured in degrees Celsius. Y is the temperature of the same resistor measured in degrees Kelvin. $\rho_{X,Y} = 1$.
- X is the gain of an electrical circuit measured in decibels. Y is the attenuation, measured in decibels, of the same circuit. $\rho_{X,Y} = -1$.
- X is the telephone number of a cellular phone. Y is the social security number of the phone's owner. $\rho_{X,Y} = 0$.

Quiz 4.7

(A) Random variables L and T given in Example 4.8 have joint PMF

$P_{L,T}(l,t)$	$t = 40$ *sec*	$t = 60$ *sec*
$l = 1$ *page*	0.15	0.1
$l = 2$ *pages*	0.30	0.2
$l = 3$ *pages*	0.15	0.1.

(4.81)

Find the following quantities.

(1) $E[L]$ *and* $\text{Var}[L]$

(2) $E[T]$ *and* $\text{Var}[T]$

(3) The correlation $r_{L,T} = E[LT]$

(4) The covariance $\text{Cov}[L,T]$

(5) The correlation coefficient $\rho_{L,T}$

(B) The joint probability density function of random variables X and Y is

$$f_{X,Y}(x,y) = \begin{cases} xy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.82}$$

Find the following quantities.

(1) $E[X]$ *and* $\text{Var}[X]$

(2) $E[Y]$ *and* $\text{Var}[Y]$

(3) The correlation $r_{X,Y} = E[XY]$

(4) The covariance $\text{Cov}[X,Y]$

(5) The correlation coefficient $\rho_{X,Y}$

4.10 Independent Random Variables

Chapter 1 presents the concept of independent events. Definition 1.7 states that events A and B are independent if and only if the probability of the intersection is the product of the individual probabilities, $P[AB] = P[A]P[B]$.

Applying the idea of independence to random variables, we say that X and Y are independent random variables if and only if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all $x \in S_X$ and all $y \in S_Y$. In terms of probability mass functions and probability density functions we have the following definition.

Definition 4.16 ***Independent Random Variables***
Random variables X and Y are ***independent*** *if and only if*

$$\textit{Discrete:} \quad P_{X,Y}(x, y) = P_X(x) P_Y(y),$$

$$\textit{Continuous:} \quad f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Because Definition 4.16 is an equality of functions, it must be true for all values of x and y. Theorem 4.22 implies that if X and Y are independent discrete random variables, then

$$P_{X|Y}(x|y) = P_X(x), \qquad P_{Y|X}(y|x) = P_Y(y). \tag{4.129}$$

Theorem 4.24 implies that if X and Y are independent continuous random variables, then

$$f_{X|Y}(x|y) = f_X(x) \qquad f_{Y|X}(y|x) = f_Y(y). \tag{4.130}$$

Example 4.23

$$f_{X,Y}(x, y) = \begin{cases} 4xy & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.131}$$

Are X and Y independent?

The marginal PDFs of X and Y are

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad f_Y(y) = \begin{cases} 2y & 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.132}$$

It is easily verified that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all pairs (x, y) and so we conclude that X and Y are independent.

Example 4.24

$$f_{U,V}(u, v) = \begin{cases} 24uv & u \ge 0, v \ge 0, u + v \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.133}$$

Are U and V independent?

Since $f_{U,V}(u, v)$ looks similar in form to $f_{X,Y}(x, y)$ in the previous example, we might suppose that U and V can also be factored into marginal PDFs $f_U(u)$ and $f_V(v)$.

Quiz 4.10

(A) *Random variables X and Y in Example 4.1 and random variables Q and G in Quiz 4.2 have joint PMFs:*

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.01	0	0
$x=1$	0.09	0.09	0
$x=2$	0	0	0.81

$P_{Q,G}(q,g)$	$g=0$	$g=1$	$g=2$	$g=3$
$q=0$	0.06	0.18	0.24	0.12
$q=1$	0.04	0.12	0.16	0.08

(1) *Are X and Y independent?* (2) *Are Q and G independent?*

(B) *Random variables X_1 and X_2 are independent and identically distributed with probability density function*

$$f_X(x) = \begin{cases} 1 - x/2 & 0 \le x \le 2, \\ 0 & \text{otherwise}. \end{cases} \tag{4.144}$$

(1) *What is the joint PDF $f_{X_1,X_2}(x_1,x_2)$?* (2) *Find the CDF of $Z = \max(X_1, X_2)$.*

4.11 Bivariate Gaussian Random Variables

The *bivariate Gaussian* disribution is a probability model for X and Y with the property that X and Y are each Gaussian random variables.

Definition 4.17 ***Bivariate Gaussian Random Variables***

*Random variables X and Y have a **bivariate Gaussian PDF** with parameters μ_1, σ_1, μ_2, σ_2, and ρ if*

$$f_{X,Y}(x,y) = \frac{\exp\left[-\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where μ_1 and μ_2 can be any real numbers, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$.

Figure 4.5 illustrates the bivariate Gaussian PDF for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and three values of ρ. When $\rho = 0$, the joint PDF has the circular symmetry of a sombrero. When $\rho = 0.9$, the joint PDF forms a ridge over the line $x = y$, and when $\rho = -0.9$ there is a ridge over the line $x = -y$. The ridge becomes increasingly steep as $\rho \to \pm 1$.

To examine mathematically the properties of the bivariate Gaussian PDF, we define

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2}, \tag{4.145}$$

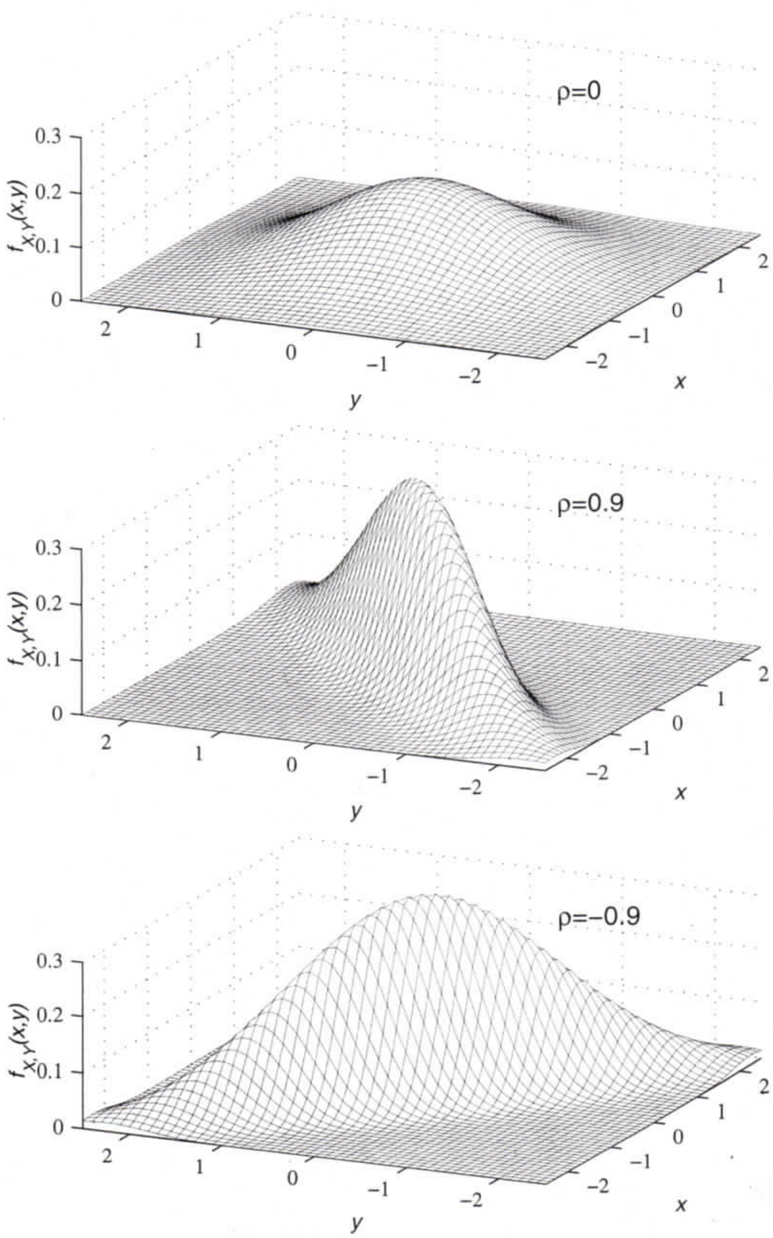

Figure 4.5 The Joint Gaussian PDF $f_{X,Y}(x, y)$ for $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$, and three values of ρ.

and manipulate the formula in Definition 4.17 to obtain the following expression for the joint Gaussian PDF:

$$f_{X,Y}(x, y) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2}. \tag{4.146}$$

Equation (4.146) expresses $f_{X,Y}(x, y)$ as the product of two Gaussian PDFs, one with parameters μ_1 and σ_1 and the other with parameters $\tilde{\mu}_2$ and $\tilde{\sigma}_2$. This formula plays a key role in the proof of the following theorem.

Theorem 4.28 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, X is the Gaussian (μ_1, σ_1) random variable and Y is the Gaussian (μ_2, σ_2) random variable:*

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \qquad f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-\mu_2)^2/2\sigma_2^2}.$$

Proof Integrating $f_{X,Y}(x, y)$ in Equation (4.146) over all y, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \tag{4.147}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2}\, dy}_{1} \tag{4.148}$$

The integral above the bracket equals 1 because it is the integral of a Gaussian PDF. The remainder of the formula is the PDF of the Gaussian (μ_1, σ_1) random variable. The same reasoning with the roles of X and Y reversed leads to the formula for $f_Y(y)$.

Given the marginal PDFs of X and Y, we use Definition 4.13 to find the conditional PDFs.

Theorem 4.29 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of Y given X is*

$$f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2},$$

where, given $X = x$, the conditional expected value and variance of Y are

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \tilde{\sigma}_2^2 = \sigma_2^2(1 - \rho^2).$$

Theorem 4.29 is the result of dividing $f_{X,Y}(x, y)$ in Equation (4.146) by $f_X(x)$ to obtain $f_{Y|X}(y|x)$. The cross sections of Figure 4.6 illustrate the conditional PDF. The figure is a graph of $f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x)$. Since X is a constant on each cross section, the cross section is a scaled picture of $f_{Y|X}(y|x)$. As Theorem 4.29 indicates, the cross section has the Gaussian bell shape. Corresponding to Theorem 4.29, the conditional PDF of X

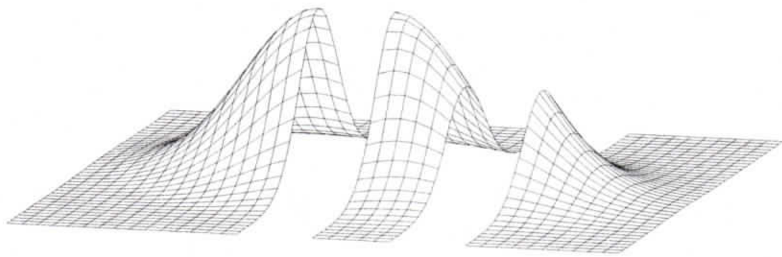

Figure 4.6 Cross-sectional view of the joint Gaussian PDF with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.9$. Theorem 4.29 confirms that the bell shape of the cross section occurs because the conditional PDF $f_{Y|X}(y|x)$ is Gaussian.

given Y is also Gaussian. This conditional PDF is found by dividing $f_{X,Y}(x, y)$ by $f_Y(y)$ to obtain $f_{X|Y}(x|y)$.

Theorem 4.30 *If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of X given Y is*

$$f_{X|Y}(x|y) = \frac{1}{\tilde{\sigma}_1\sqrt{2\pi}} e^{-(x-\tilde{\mu}_1(y))^2/2\tilde{\sigma}_1^2},$$

where, given $Y = y$, the conditional expected value and variance of X are

$$\tilde{\mu}_1(y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2) \qquad \tilde{\sigma}_1^2 = \sigma_1^2(1 - \rho^2).$$

The next theorem identifies ρ in Definition 4.17 as the correlation coefficient of X and Y, $\rho_{X,Y}$.

Theorem 4.31 *Bivariate Gaussian random variables X and Y in Definition 4.17 have correlation coefficient*

$$\rho_{X,Y} = \rho.$$

Proof Substituting μ_1, σ_1, μ_2, and σ_2 for μ_X, σ_X, μ_Y, and σ_Y in Definition 4.4 and Definition 4.8, we have

$$\rho_{X,Y} = \frac{E\left[(X - \mu_1)(Y - \mu_2)\right]}{\sigma_1\sigma_2}. \tag{4.149}$$

Theo

Qui

To evaluate this expected value, we use the substitution $f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$ in the double integral of Theorem 4.12. The result can be expressed as

$$\rho_{X,Y} = \frac{1}{\sigma_1\sigma_2}\int_{-\infty}^{\infty}(x-\mu_1)\left(\int_{-\infty}^{\infty}(y-\mu_2)\, f_{Y|X}(y|x)\, dy\right) f_X(x)\, dx \tag{4.150}$$

$$= \frac{1}{\sigma_1\sigma_2}\int_{-\infty}^{\infty}(x-\mu_1)\, E\left[Y-\mu_2|X=x\right] f_X(x)\, dx \tag{4.151}$$

Because $E[Y|X=x] = \tilde{\mu}_2(x)$ in Theorem 4.29, it follows that

$$E\left[Y-\mu_2|X=x\right] = \tilde{\mu}_2(x) - \mu_2 = \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \tag{4.152}$$

Therefore,

$$\rho_{X,Y} = \frac{\rho}{\sigma_1^2}\int_{-\infty}^{\infty}(x-\mu_1)^2 f_X(x)\, dx = \rho, \tag{4.153}$$

because the integral in the final expression is $\text{Var}[X] = \sigma_1^2$.

From Theorem 4.31, we observe that if X and Y are uncorrelated, then $\rho = 0$ and, from Theorems 4.29 and 4.30, $f_{Y|X}(y|x) = f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$. Thus we have the following theorem.

Theorem 4.32 *Bivariate Gaussian random variables X and Y are uncorrelated if and only if they are independent.*

Theorem 4.31 identifies the parameter ρ in the bivariate gaussian PDF as the correlation coefficient $\rho_{X,Y}$ of bivariate Gaussian random variables X and Y. Theorem 4.17 states that for any pair of random variables, $|\rho_{X,Y}| < 1$, which explains the restriction $|\rho| < 1$ in Definition 4.17. Introducing this inequality to the formulas for conditional variance in Theorem 4.29 and Theorem 4.30 leads to the following inequalities:

$$\text{Var}\left[Y|X=x\right] = \sigma_2^2(1-\rho^2) \le \sigma_2^2, \tag{4.154}$$

$$\text{Var}\left[X|Y=y\right] = \sigma_1^2(1-\rho^2) \le \sigma_1^2. \tag{4.155}$$

These formulas state that for $\rho \neq 0$, learning the value of one of the random variables leads to a model of the other random variable with reduced variance. This suggests that learning the value of Y reduces our uncertainty regarding X.

Quiz 4.11 *Let X and Y be jointly Gaussian $(0, 1)$ random variables with correlation coefficient $1/2$.*

(1) What is the joint PDF of X and Y?

(2) What is the conditional PDF of X given $Y = 2$?