CHAPTER 5

Multivariate Probability Distributions

5.1 Introduction

The intersection of two or more events is frequently of interest to an experimenter. For example, a gambler playing blackjack is interested in the event of drawing both an ace and a face card from a 52-card deck. A biologist, observing the number of animals surviving in a litter, is concerned about the intersection of these events:

A: The litter contains n animals.
B: y animals survive.

Similarly, observing both the height and the weight of an individual represents the intersection of a specific pair of events associated with height–weight measurements.

Most important to statisticians are intersections that occur in the course of sampling. Suppose that $Y_1, Y_2, \ldots, Y_n$ denote the outcomes of n successive trials of an experiment. For example, this sequence could represent the weights of n people or the measurements of n physical characteristics for a single person. A specific set of outcomes, or sample measurements, may be expressed in terms of the intersection of the n events $(Y_1 = y_1), (Y_2 = y_2), \ldots, (Y_n = y_n)$, which we will denote as $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$, or, more compactly, as $(y_1, y_2, \ldots, y_n)$. Calculation of the probability of this intersection is essential in making inferences about the population from which the sample was drawn and is a major reason for studying multivariate probability distributions.

5.2 Bivariate and Multivariate Probability Distributions

Many random variables can be defined over the same sample space. For example, consider the experiment of tossing a pair of dice. The sample space contains 36 sample points, corresponding to the $mn = (6)(6) = 36$ ways in which numbers may appear on the faces of the dice. Any one of the following random variables could be defined over the sample space and might be of interest to the experimenter:

Y_1: The number of dots appearing on die 1.
Y_2: The number of dots appearing on die 2.
Y_3: The sum of the number of dots on the dice.
Y_4: The product of the number of dots appearing on the dice.

The 36 sample points associated with the experiment are equiprobable and correspond to the 36 numerical events (y_1, y_2). Thus, throwing a pair of 1s is the simple event (1, 1). Throwing a 2 on die 1 and a 3 on die 2 is the simple event (2, 3). Because all pairs (y_1, y_2) occur with the same relative frequency, we assign probability $1/36$ to each sample point. For this simple example, the intersection (y_1, y_2) contains at most one sample point. Hence, the bivariate probability function is

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) = 1/36, \qquad y_1 = 1, 2, \ldots, 6, y_2 = 1, 2, \ldots, 6.$$

A graph of the bivariate probability function for the die-tossing experiment is shown in Figure 5.1. Notice that a nonzero probability is assigned to a point (y_1, y_2) in the plane if and only if $y_1 = 1, 2, \ldots, 6$ and $y_2 = 1, 2, \ldots, 6$. Thus, exactly 36 points in the plane are assigned nonzero probabilities. Further, the probabilities are assigned in such a way that the sum of the nonzero probabilities is equal to 1. In Figure 5.1 the points assigned nonzero probabilities are represented in the (y_1, y_2) plane, whereas the probabilities associated with these points are given by the lengths of the lines above them. Figure 5.1 may be viewed as a theoretical, three-dimensional relative frequency histogram for the pairs of observations (y_1, y_2). As in the single-variable discrete case, the theoretical histogram provides a model for the sample histogram that would be obtained if the die-tossing experiment were repeated a large number of times.

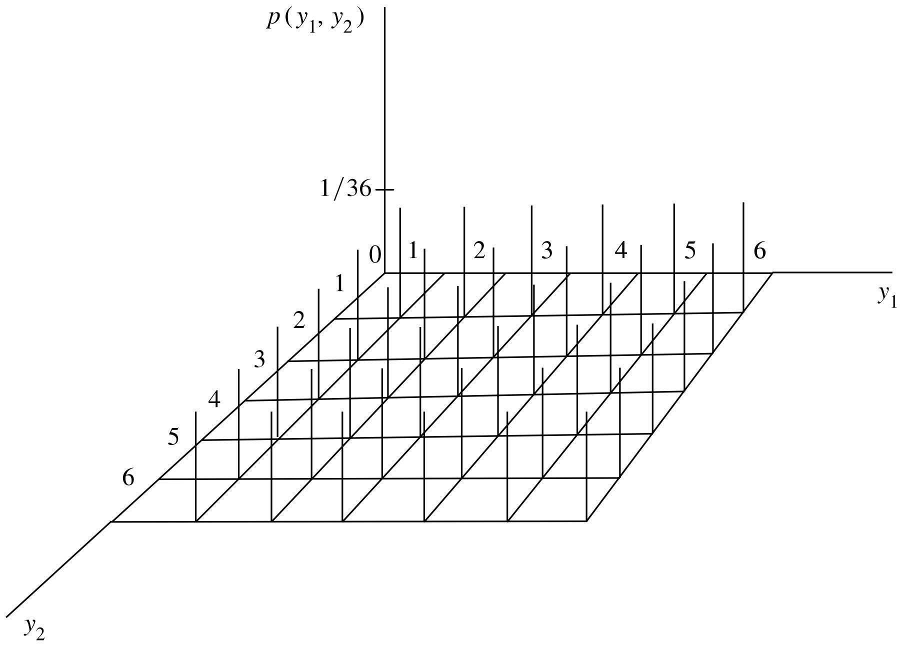


FIGURE **5.1** Bivariate probability function; y_1 = number of dots on die 1, y_2 = number of dots on die 2

DEFINITION **5.1**

Let Y_1 and Y_2 be discrete random variables. The *joint* (or bivariate) *probability function* for Y_1 and Y_2 is given by

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

In the single-variable case discussed in Chapter 3, we saw that the probability function for a discrete random variable Y assigns nonzero probabilities to a finite or countable number of distinct values of Y in such a way that the sum of the probabilities is equal to 1. Similarly, in the bivariate case the joint probability function $p(y_1, y_2)$ assigns nonzero probabilities to only a finite or countable number of pairs of values (y_1, y_2). Further, the nonzero probabilities must sum to 1.

THEOREM **5.1**

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$, where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities.

As in the univariate discrete case, the joint probability function for discrete random variables is sometimes called the *joint probability mass function* because it specifies the probability (mass) associated with each of the possible pairs of values for the random variables. Once the joint probability function has been determined for discrete random variables Y_1 and Y_2, calculating joint probabilities involving Y_1 and Y_2 is straightforward. For the die-tossing experiment, $P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2)$ is

$$P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2) = p(2, 1) + p(2, 2) + p(3, 1) + p(3, 2)$$
$$= 4/36 = 1/9.$$

EXAMPLE **5.1** A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let Y_1 denote the number of customers who choose counter 1 and Y_2, the number who select counter 2. Find the joint probability function of Y_1 and Y_2.

Solution We might proceed with the derivation in many ways. The most direct is to consider the sample space associated with the experiment. Let the pair $\{i, j\}$ denote the simple event that the first customer chose counter i and the second customer chose counter j, where $i, j = 1, 2$, and 3. Using the mn rule, the sample space consists of $3 \times 3 = 9$ sample points. Under the assumptions given earlier, each sample point is equally likely and has probability $1/9$. The sample space associated with the experiment is

$$S = [\{1,\ 1\},\ \{1,\ 2\},\ \{1,\ 3\},\ \{2,\ 1\},\ \{2,\ 2\},\ \{2,\ 3\},\ \{3,\ 1\},\ \{3,\ 2\},\ \{3,\ 3\}].$$

Notice that sample point $\{1,\ 1\}$ is the only sample point corresponding to $(Y_1 = 2, Y_2 = 0)$ and hence $P(Y_1 = 2, Y_2 = 0) = 1/9$. Similarly, $P(Y_1 = 1, Y_2 = 1) = P(\{1, 2\} \text{ or } \{2, 1\}) = 2/9$. Table 5.1 contains the probabilities associated with each possible pair of values for Y_1 and Y_2—that is, the joint probability function for Y_1 and Y_2. As always, the results of Theorem 5.1 hold for this example.

Table 5.1 Probability function for Y_1 and Y_2, Example 5.1

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

■

As in the case of univariate random variables, the distinction between jointly discrete and jointly continuous random variables may be characterized in terms of their (joint) distribution functions.

DEFINITION **5.2** For any random variables Y_1 and Y_2, the joint (bivariate) distribution function $F(y_1, y_2)$ is

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

For two discrete variables Y_1 and Y_2, $F(y_1, y_2)$ is given by

$$F(y_1, y_2) = \sum_{t_1 \le y_1} \sum_{t_2 \le y_2} p(t_1, t_2).$$

For the die-tossing experiment,

$$\begin{aligned} F(2, 3) &= P(Y_1 \le 2, Y_2 \le 3) \\ &= p(1, 1) + p(1, 2) + p(1, 3) + p(2, 1) + p(2, 2) + p(2, 3). \end{aligned}$$

Because $p(y_1, y_2) = 1/36$ for all pairs of values of y_1 and y_2 under consideration, $F(2, 3) = 6/36 = 1/6$.

EXAMPLE 5.2 Consider the random variables Y_1 and Y_2 of Example 5.1. Find $F(-1, 2)$, $F(1.5, 2)$, and $F(5, 7)$.

Solution Using the results in Table 5.1, we see that

$$F(-1, 2) = P(Y_1 \le -1, Y_2 \le 2) = P(\emptyset) = 0.$$

Further,

$$\begin{aligned} F(1.5, 2) &= P(Y_1 \le 1.5,\ Y_2 \le 2) \\ &= p(0, 0) + p(0, 1) + p(0, 2) + p(1, 0) + p(1, 1) + p(1, 2) = 8/9. \end{aligned}$$

Similarly,

$$F(5, 7) = P(Y_1 \le 5,\ Y_2 \le 7) = 1.$$

Notice that $F(y_1, y_2) = 1$ for all y_1, y_2 such that $\min\{y_1, y_2\} \ge 2$. Also, $F(y_1, y_2) = 0$ if $\min\{y_1, y_2) < 0$. ■

Two random variables are said to be jointly continuous if their joint distribution function $F(y_1, y_2)$ is continuous in both arguments.

DEFINITION 5.3 Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a nonnegative function $f(y_1, y_2)$, such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1,$$

for all $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$, then Y_1 and Y_2 are said to be *jointly continuous random variables*. The function $f(y_1, y_2)$ is called the *joint probability density function*.

Bivariate cumulative distribution functions satisfy a set of properties similar to those specified for univariate cumulative distribution functions.

THEOREM **5.2**

If Y_1 and Y_2 are random variables with joint distribution function $F(y_1, y_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. If $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0.$$

Part 3 follows because

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2)$$
$$= P(y_1 < Y_1 \leq y_1^*, y_2 < Y_2 \leq y_2^*) \geq 0.$$

Notice that $F(\infty, \infty) \equiv \lim_{y_1 \to \infty} \lim_{y_2 \to \infty} F(y_1, y_2) = 1$ implies that the joint density function $f(y_1, y_2)$ must be such that the integral of $f(y_1, y_2)$ over all values of (y_1, y_2) is 1.

THEOREM **5.2**

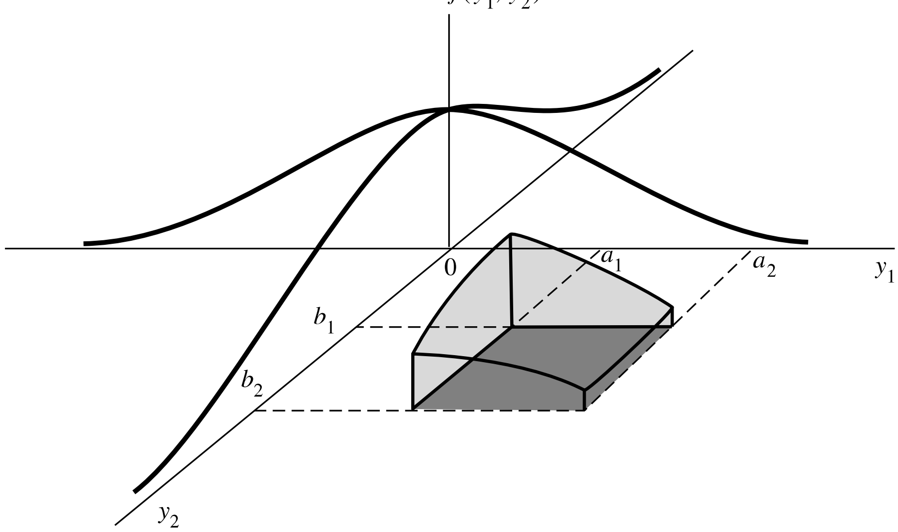
If Y_1 and Y_2 are jointly continuous random variables with a joint density function given by $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = 1$.

As in the univariate continuous case discussed in Chapter 4, the joint density function may be intuitively interpreted as a model for the joint relative frequency histogram for Y_1 and Y_2.

For the univariate continuous case, areas under the probability density over an interval correspond to probabilities. Similarly, the bivariate probability density function $f(y_1, y_2)$ traces a probability density surface over the (y_1, y_2) plane (Figure 5.2).

FIGURE **5.2**
A bivariate density function $f(y_1, y_2)$



Volumes under this surface correspond to probabilities. Thus, $P(a_1 \le Y_1 \le a_2, b_1 \le Y_2 \le b_2)$ is the shaded volume shown in Figure 5.2 and is equal to

$$\int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2)\, dy_1\, dy_2.$$

EXAMPLE **5.3** Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions within the unit square and of equal area are considered, the particle is equally likely to be in either region. Let Y_1 and Y_2 denote the coordinates of the particle's location. A reasonable model for the relative frequency histogram for Y_1 and Y_2 is the bivariate analogue of the univariate uniform density function:

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a Sketch the probability density surface.
b Find $F(.2, .4)$.
c Find $P(.1 \le Y_1 \le .3, 0 \le Y_2 \le .5)$.

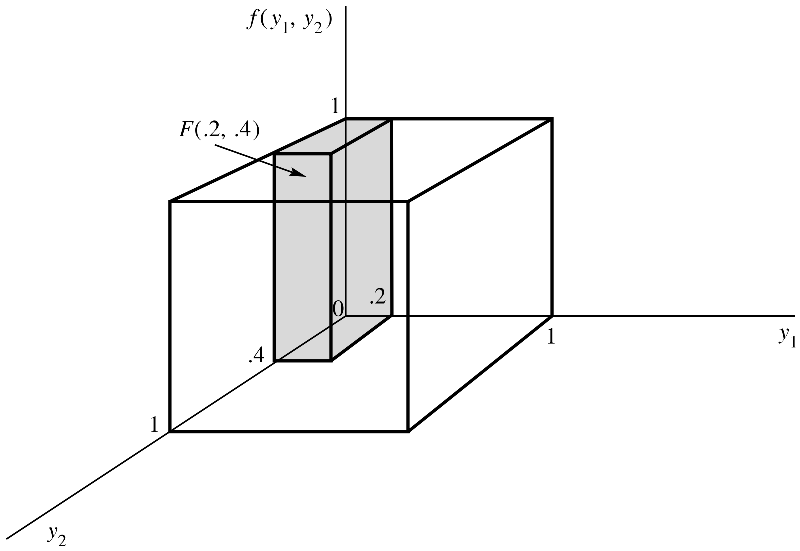
Solution **a** The sketch is shown in Figure 5.3.

b

$$\begin{aligned} F(.2, .4) &= \int_{-\infty}^{.4} \int_{-\infty}^{.2} f(y_1, y_2)\, dy_1\, dy_2 \\ &= \int_{0}^{.4} \int_{0}^{.2} (1)\, dy_1\, dy_2 \\ &= \int_{0}^{.4} \left(y_1 \Big]_0^{.2} \right) dy_2 = \int_0^{.4} .2\, dy_2 = .08. \end{aligned}$$

The probability $F(.2, .4)$ corresponds to the volume under $f(y_1, y_2) = 1$, which is shaded in Figure 5.3. As geometric considerations indicate, the desired probability (volume) is equal to .08, which we obtained through integration at the beginning of this part.

FIGURE **5.3** Geometric representation of $f(y_1, y_2)$, Example 5.3



c
$$P(.1 \le Y_1 \le .3, 0 \le Y_2 \le .5) = \int_0^{.5}\int_{.1}^{.3} f(y_1, y_2)\, dy_1\, dy_2$$
$$= \int_0^{.5}\int_{.1}^{.3} 1\, dy_1\, dy_2 = .10.$$

This probability corresponds to the volume under the density function $f(y_1, y_2) = 1$ that is above the region $.1 \le y_1 \le .3, 0 \le y_2 \le .5$. Like the solution in part (b), the current solution can be obtained by using elementary geometric concepts. The density or height of the surface is equal to 1, and hence the desired probability (volume) is

$$P(.1 \le Y_1 \le .3, 0 \le Y_2 \le .5) = (.2)(.5)(1) = .10.$$ ■

A slightly more complicated bivariate model is illustrated in the following example.

EXAMPLE **5.4** Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables take on values between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose that the joint density function for Y_1 and Y_2 is given by

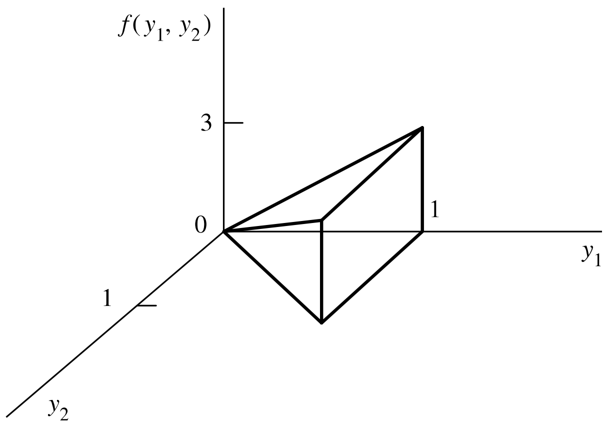
$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

A sketch of this function is given in Figure 5.4.

Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

Solution We want to find $P(0 \le Y_1 \le .5, Y_2 > .25)$. For any continuous random variable, the probability of observing a value in a region is the volume under the density function above the region of interest. The density function $f(y_1, y_2)$ is positive only in the

FIGURE **5.4** The joint density function for Example 5.4



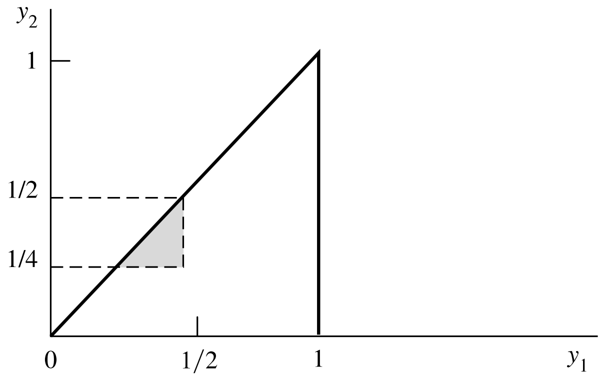


FIGURE 5.5 Region of integration for Example 5.4

large triangular portion of the (y_1, y_2) plane shown in Figure 5.5. We are interested only in values of y_1 and y_2 such that $0 \leq y_1 \leq .5$ and $y_2 > .25$. The intersection of this region and the region where the density function is positive is given by the small (shaded) triangle in Figure 5.5. Consequently, the probability we desire is the volume under the density function of Figure 5.4 above the shaded region in the (y_1, y_2) plane shown in Figure 5.5.

Thus, we have

$$
\begin{aligned}
P(0 \leq Y_1 \leq .5, .25 \leq Y_2) &= \int_{1/4}^{1/2} \int_{1/4}^{y_1} 3y_1 \, dy_2 \, dy_1 \\
&= \int_{1/4}^{1/2} 3y_1 \left(y_2 \Big]_{1/4}^{y_1} \right) dy_1 \\
&= \int_{1/4}^{1/2} 3y_1(y_1 - 1/4) \, dy_1 \\
&= \left[y_1^3 - (3/8)y_1^2 \right] \Big]_{1/4}^{1/2} \\
&= [(1/8) - (3/8)(1/4)] - [(1/64) - (3/8)(1/16)] \\
&= 5/128.
\end{aligned}
$$

■

Calculating the probability specified in Example 5.4 involved integrating the joint density function for Y_1 and Y_2 over the appropriate region. The specification of the limits of integration was made easier by sketching the region of integration in Figure 5.5. This approach, sketching the appropriate region of integration, often facilitates setting up the appropriate integral.

The methods discussed in this section can be used to calculate the probability of the intersection of two events $(Y_1 = y_1, Y_2 = y_2)$. In a like manner, we can define a probability function (or probability density function) for the intersection of n events $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$. The joint probability function corresponding to the discrete case is given by

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n).$$

The joint density function of $Y_1, Y_2, \ldots, Y_n$ is given by $f(y_1, y_2, \ldots, y_n)$. As in the bivariate case, these functions provide models for the joint relative frequency distributions of the populations of joint observations $(y_1, y_2, \ldots, y_n)$ for the discrete case and the continuous case, respectively. In the continuous case,

$$P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) = F(y_1, \ldots, y_n)$$
$$= \int_{-\infty}^{y_1}\int_{-\infty}^{y_2}\cdots\int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n)dt_n \ldots dt_1$$

for every set of real numbers $(y_1, y_2, \ldots, y_n)$. Multivariate distribution functions defined by this equality satisfy properties similar to those specified for the bivariate case.

Exercises

5.1 Contracts for two construction jobs are randomly assigned to one or more of three firms, A, B, and C. Let Y_1 denote the number of contracts assigned to firm A and Y_2 the number of contracts assigned to firm B. Recall that each firm can receive 0, 1, or 2 contracts.

a Find the joint probability function for Y_1 and Y_2.

b Find $F(1, 0)$.

5.2 Three balanced coins are tossed independently. One of the variables of interest is Y_1, the number of heads. Let Y_2 denote the amount of money won on a side bet in the following manner. If the first head occurs on the first toss, you win \$1. If the first head occurs on toss 2 or on toss 3 you win \$2 or \$3, respectively. If no heads appear, you lose \$1 (that is, win −\$1).

a Find the joint probability function for Y_1 and Y_2.

b What is the probability that fewer than three heads will occur and you will win \$1 or less? [That is, find $F(2, 1)$.]

5.3 Of nine executives in a business firm, four are married, three have never married, and two are divorced. Three of the executives are to be selected for promotion. Let Y_1 denote the number of married executives and Y_2 denote the number of never-married executives among the three selected for promotion. Assuming that the three are randomly selected from the nine available, find the joint probability function of Y_1 and Y_2.

5.4 Given here is the joint probability function associated with data obtained in a study of automobile accidents in which a child (under age 5 years) was in the car and at least one fatality occurred. Specifically, the study focused on whether or not the child survived and what type of seatbelt (if any) he or she used. Define

$$Y_1 = \begin{cases} 0, & \text{if the child survived,} \\ 1, & \text{if not,} \end{cases} \quad \text{and} \quad Y_2 = \begin{cases} 0, & \text{if no belt used,} \\ 1, & \text{if adult belt used,} \\ 2, & \text{if car-seat belt used.} \end{cases}$$

Notice that Y_1 is the number of fatalities per child and, since children's car seats usually utilize two belts, Y_2 is the number of seatbelts in use at the time of the accident.

	y_1		
y_2	0	1	Total
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
Total	.76	.24	1.00

a Verify that the preceding probability function satisfies Theorem 5.1.

b Find $F(1, 2)$. What is the interpretation of this value?

5.5 Refer to Example 5.4. The joint density of Y_1, the proportion of the capacity of the tank that is stocked at the beginning of the week, and Y_2, the proportion of the capacity sold during the week, is given by

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $F(1/2, 1/3) = P(Y_1 \le 1/2, Y_2 \le 1/3)$.

b Find $P(Y_2 \le Y_1/2)$, the probability that the amount sold is less than half the amount purchased.

5.6 Refer to Example 5.3. If a radioactive particle is randomly located in a square of unit length, a reasonable model for the joint density function for Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a What is $P(Y_1 - Y_2 > .5)$?

b What is $P(Y_1 Y_2 < .5)$?

5.7 Let Y_1 and Y_2 have joint density function

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

a What is $P(Y_1 < 1, Y_2 > 5)$?

b What is $P(Y_1 + Y_2 < 3)$?

5.8 Let Y_1 and Y_2 have the joint probability density function given by

$$f(y_1, y_2) = \begin{cases} ky_1y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the value of k that makes this a probability density function.

b Find the joint distribution function for Y_1 and Y_2.

c Find $P(Y_1 \le 1/2, Y_2 \le 3/4)$.

5.9 Let Y_1 and Y_2 have the joint probability density function given by

$$f(y_1, y_2) = \begin{cases} k(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the value of k that makes this a probability density function.

b Find $P(Y_1 \le 3/4, Y_2 \ge 1/2)$.

5.10 An environmental engineer measures the amount (by weight) of particulate pollution in air samples of a certain volume collected over two smokestacks at a coal-operated power plant. One of the stacks is equipped with a cleaning device. Let Y_1 denote the amount of pollutant per sample collected above the stack that has no cleaning device and let Y_2 denote the amount of pollutant per sample collected above the stack that is equipped with the cleaning device.