5 Joint Probability Distributions

CHAPTER OUTLINE

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Use joint probability mass functions and joint probability density functions to calculate probabilities
2. Calculate marginal and conditional probability distributions from joint probability distributions
3. Use the multinomial distribution to determine probabilities
4. Interpret and calculate covariances and correlations between random variables
5. Understand properties of a bivariate normal distribution and be able to draw contour plots for the probability density function
6. Calculate means and variance for linear combinations of random variables and calculate probabilities for linear combinations of normally distributed random variables
7. Determine the distribution of a general function of a random variable

5-1 Two Discrete Random Variables

Example 5-1

In the development of a new receiver for the transmission of digital information, each received bit is rated as *acceptable, suspect,* or *unacceptable,* depending on the quality of the received signal, with probabilities 0.9, 0.08, and 0.02, respectively. Assume that the ratings of each bit are independent.

In the first four bits transmitted, let

X denote the number of acceptable bits

Y denote the number of suspect bits

Then, the distribution of X is binomial with $n = 4$ and $p = 0.9$, and the distribution of Y is binomial with $n = 4$ and $p = 0.08$. However, because only four bits are being rated, the possible values of X and Y are restricted to the points shown in the graph in Fig. 5-1. Although the possible values of X are 0, 1, 2, 3, or 4, if $y = 3$, $x = 0$ or 1. By specifying the probability of each of the points in Fig. 5-1, we specify the joint probability distribution of X and Y. Similarly to an individual random variable, we define the range of the random variables (X, Y) to be the set of points (x, y) in two-dimensional space for which the probability that $X = x$ and $Y = y$ is positive.

5-1 Two Discrete Random Variables

	x = number of bars of signal strength			
y = number of times city name is stated	1	2	3	
4	0.15	0.1	0.05	
3	0.02	0.1	0.05	
2	0.02	0.03	0.2	
1	0.01	0.02	0.25	

Figure 5-1 Joint probability distribution of X and Y in Example 5-1.

5-1 Two Discrete Random Variables

5-1.1 Joint Probability Distributions

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

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

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

5-1 Two Discrete Random Variables

5-1.2 Marginal Probability Distributions

• The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.

• In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

5-1 Two Discrete Random Variables

Example 5-2

The joint probability distribution of X and Y in Fig. 5-1 can be used to find the marginal probability distribution of X. For example,

$$\begin{aligned} P(X = 3) &= P(X = 3, Y = 0) + P(X = 3, Y = 1) \\ &= 0.0583 + 0.2333 = 0.292 \end{aligned}$$

As expected, this probability matches the result obtained from the binomial probability distribution for X; that is, $P(X = 3) = \binom{4}{3}0.9^3 0.1^1 = 0.292$. The marginal probability distribution for X is found by summing the probabilities in each column, whereas the marginal probability distribution for Y is found by summing the probabilities in each row. The results are shown in Fig. 5-2.

5-1 Two Discrete Random Variables

	x = number of bars of signal strength				
y = number of times city name is stated	1	2	3		Marginal probability distribution of Y
4	0.15	0.1	0.05		0.3
3	0.02	0.1	0.05		0.17
2	0.02	0.03	0.2		0.25
1	0.01	0.02	0.25		0.28
	0.2	0.25	0.55		
	Marginal probability distribution of X				

Figure 5-2 Marginal probability distributions of X and Y from Figure 5-1.

5-1 Two Discrete Random Variables

Definition: Marginal Probability Mass Functions

If X and Y are discrete random variables with joint probability mass function $f_{XY}(x, y)$, then the **marginal probability mass functions** of X and Y are

$$f_X(x) = P(X = x) = \sum_y f_{XY}(x, y) \quad \text{and} \quad f_Y(y) = P(Y = y) = \sum_x f_{XY}(x, y) \tag{5-2}$$

where the first sum is over all points in the range of (X, Y) for which $X = x$ and the second sum is over all points in the range of (X, Y) for which $Y = y$

5-1 Two Discrete Random Variables

5-1.3 Conditional Probability Distributions

Given discrete random variables X and Y with joint probability mass function $f_{XY}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_X(x) \qquad \text{for } f_X(x) > 0 \qquad (5\text{-}3)$$

5-1 Two Discrete Random Variables

5-1.3 Conditional Probability Distributions

Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\sum_{y} f_{Y|x}(y) = 1$

(3) $P(Y = y | X = x) = f_{Y|x}(y)$ (5-4)

5-1 Two Discrete Random Variables

Definition: Conditional Mean and Variance

The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \sum_y y f_{Y|x}(y) \tag{5-5}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \sum_y (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_y y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

Example 5-4

For the joint probability distribution in Fig. 5-1, $f_{Y|x}(y)$ is found by dividing each $f_{XY}(x, y)$ by $f_X(x)$. Here, $f_X(x)$ is simply the sum of the probabilities in each column of Fig. 5-1. The function $f_{Y|x}(y)$ is shown in Fig. 5-3. In Fig. 5-3, each column sums to one because it is a probability distribution.

	x = number of bars of signal strength		
y = number of times city name is stated	1	2	3
4	0.750	0.400	0.091
3	0.100	0.400	0.091
2	0.100	0.120	0.364
1	0.050	0.080	0.454

Figure 5-3 Conditional probability distributions of Y given X, $f_{Y|x}(y)$ in Example 5-6.

5-1 Two Discrete Random Variables

5-1.4 Independence

Example 5-6

In a plastic molding operation, each part is classified as to whether it conforms to color and length specifications. Define the random variable X and Y as

$$X = \begin{cases} 1 & \text{if the part conforms to color specifications} \\ 0 & \text{otherwise} \end{cases}$$

$$Y = \begin{cases} 1 & \text{if the part conforms to length specifications} \\ 0 & \text{otherwise} \end{cases}$$

Assume the joint probability distribution of X and Y is defined by $f_{XY}(x, y)$ in Fig. 5-4(a). The marginal probability distributions of X and Y are also shown in Fig. 5-4(a). Note that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$. The conditional probability mass function $f_{Y|x}(y)$ is shown in Fig. 5-4(b). Notice that for any x, $f_{Y|x}(y) = f_Y(y)$. That is, knowledge of whether or not the part meets color specifications does not change the probability that it meets length specifications.

Example 5-8

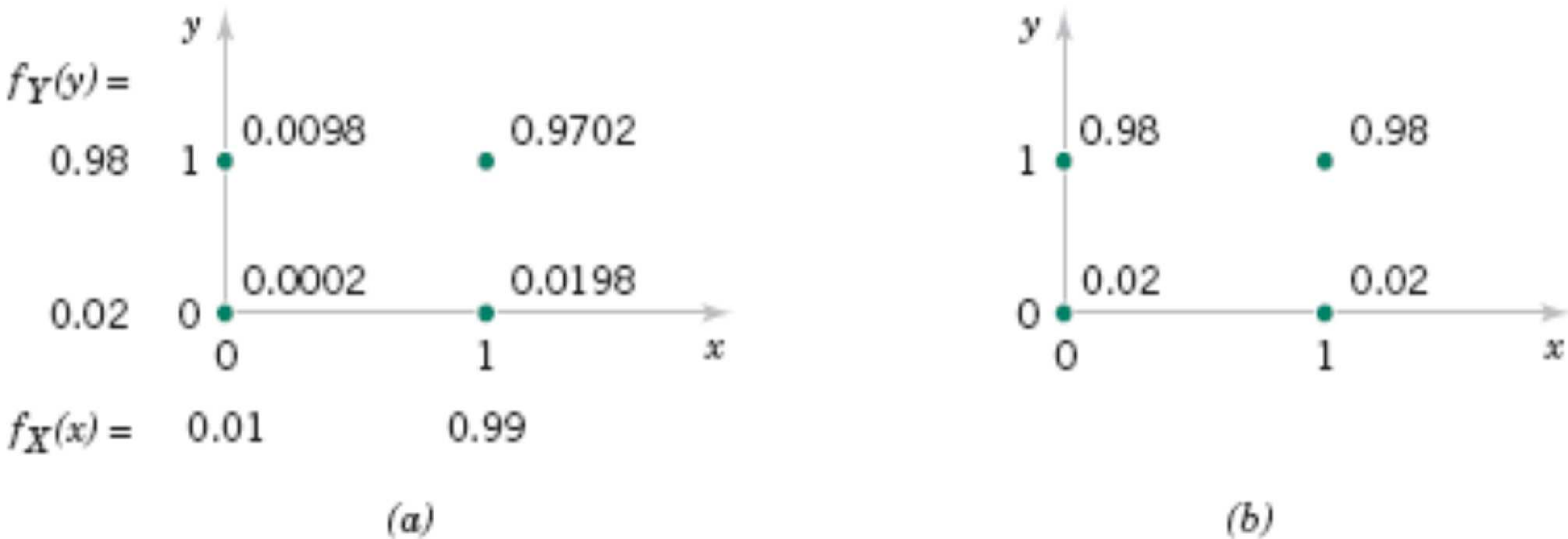

Figure 5-4 (a)Joint and marginal probability distributions of X and Y in Example 5-8. (b) Conditional probability distribution of Y given $X = x$ in Example 5-8.

5-1 Two Discrete Random Variables

5-1.4 Independence

For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-6)

5-1 Two Discrete Random Variables

5-1.5 Multiple Discrete Random Variables

Definition: Joint Probability Mass Function

The **joint probability mass function** of $X_1, X_2, \ldots, X_p$ is

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \qquad (5\text{-}7)$$

for all points $(x_1, x_2, \ldots, x_p)$ in the range of $X_1, X_2, \ldots, X_p$.

A marginal probability distribution is a simple extension of the result for two random variables.

5-1 Two Discrete Random Variables

5-1.5 Multiple Discrete Random Variables

Definition: Marginal Probability Mass Function

If $X_1, X_2, X_3, \ldots, X_p$ are discrete random variables with joint probability mass function $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$, the **marginal probability mass function** of any X_i is

$$f_{X_i}(x_i) = P(X_i = x_i) = \sum f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \qquad (5\text{-}8)$$

where the sum is over the points in the range of $(X_1, X_2, \ldots, X_p)$ for which $X_i = x_i$.

5-1 Two Discrete Random Variables

Example 5-8

Points that have positive probability in the joint probability distribution of three random variables X_1, X_2, X_3 are shown in Fig. 5-5. The range is the nonnegative integers with $x_1 + x_2 + x_3 = 3$. The marginal probability distribution of X_2 is found as follows.

$$P(X_2 = 0) = f_{X_1X_2X_3}(3, 0, 0) + f_{X_1X_2X_3}(0, 0, 3) + f_{X_1X_2X_3}(1, 0, 2) + f_{X_1X_2X_3}(2, 0, 1)$$
$$P(X_2 = 1) = f_{X_1X_2X_3}(2, 1, 0) + f_{X_1X_2X_3}(0, 1, 2) + f_{X_1X_2X_3}(1, 1, 1)$$
$$P(X_2 = 2) = f_{X_1X_2X_3}(1, 2, 0) + f_{X_1X_2X_3}(0, 2, 1)$$
$$P(X_2 = 3) = f_{X_1X_2X_3}(0, 3, 0)$$

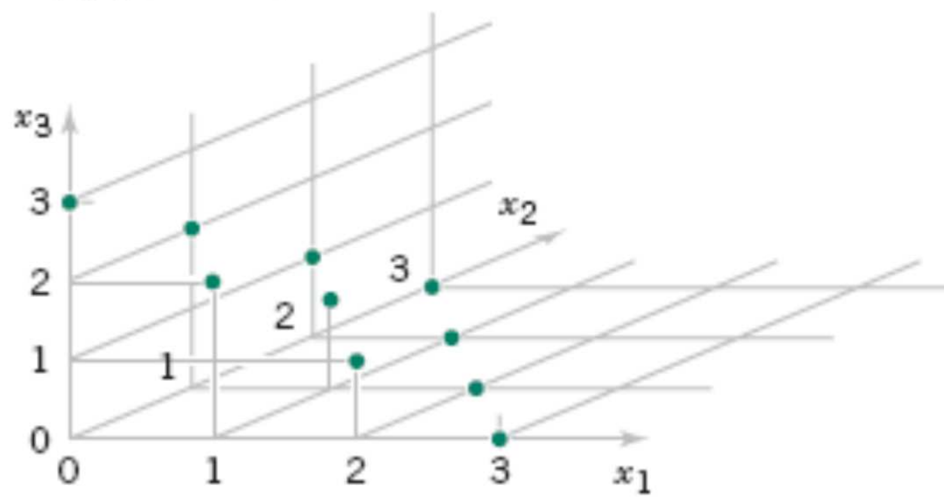

Figure 5-5 Joint probability distribution of X_1, X_2, and X_3.

5-1 Two Discrete Random Variables

5-1.5 Multiple Discrete Random Variables

Mean and Variance from Joint Probability

$$E(X_i) = \sum x_i f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$$

and

$$V(X_i) = \sum (x_i - \mu_{X_i})^2 f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) \qquad (5\text{-}9)$$

where the sum is over all points in the range of $X_1, X_2, \ldots, X_p$.

5-1 Two Discrete Random Variables

5-1.5 Multiple Discrete Random Variables

Distribution of a Subset of Random Variables

If $X_1, X_2, X_3, \ldots, X_p$ are discrete random variables with joint probability mass function $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, the **joint probability mass function** of $X_1, X_2, \ldots, X_k$, $k < p$, is

$$f_{X_1X_2\ldots X_k}(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$$
$$= \sum_{R_{x_1x_2\ldots x_k}} P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) \qquad (5\text{-}10)$$

where the sum is over all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_1 = x_1$, $X_2 = x_2, \ldots, X_k = x_k$.

5-1 Two Discrete Random Variables

5-1.5 Multiple Discrete Random Variables

Conditional Probability Distributions

Discrete variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = f_{X_1}(x_1)\, f_{X_2}(x_2) \ldots f_{X_p}(x_p) \qquad \text{(5-11)}$$

for *all* $x_1, x_2, \ldots, x_p$.

5-1 Two Discrete Random Variables

5-1.6 Multinomial Probability Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, … , class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, … , class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \qquad (5\text{-}12)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

5-1 Two Discrete Random Variables

5-1.6 Multinomial Probability Distribution

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \tag{5-13}$$