

Developing Discrete Probability Distributions

5.2

The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable. For a discrete random variable x , a **probability function**, denoted by $f(x)$, provides the probability for each value of the random variable. As such, you might suppose that the classical, subjective, and relative frequency methods of assigning probabilities introduced in Chapter 4 would be useful in developing discrete probability distributions. They are, and in this section we show how. Application of this methodology leads to what we call tabular discrete probability distributions, that is, probability distributions that are presented in a table.

The classical method of assigning probabilities to values of a random variable is applicable when the experimental outcomes generate values of the random variable that are equally likely. For instance, consider the experiment of rolling a die and observing the number on the upward face. It must be one of the numbers 1, 2, 3, 4, 5, or 6, and each of these outcomes is equally likely. Thus, if we let x = number obtained on one roll of a die and $f(x)$ = the probability of x , the probability distribution of x is given in Table 5.3.

The subjective method of assigning probabilities can also lead to a table of values of the random variable together with the associated probabilities. With the subjective method the individual developing the probability distribution uses their best judgment to assign each probability. So, unlike probability distributions developed using the classical method, different people can be expected to obtain different probability distributions.

The relative frequency method of assigning probabilities to values of a random variable is applicable when reasonably large amounts of data are available. We then treat the data as if they were the population and use the relative frequency method to assign probabilities to the experimental outcomes. The use of the relative frequency method to develop discrete probability distributions leads to what is called an **empirical discrete distribution**. With the large amounts of data available today (e.g., scanner data, credit card data), this type of probability distribution is becoming more widely used in practice. Let us illustrate by considering the sale of automobiles at a dealership.

We will use the relative frequency method to develop a probability distribution for the number of cars sold per day at DiCarlo Motors in Saratoga, New York. Over the past 300 days, DiCarlo has experienced 54 days with no automobiles sold, 117 days with one automobile sold, 72 days with two automobiles sold, 42 days with three automobiles sold, 12 days with four automobiles sold, and 3 days with five automobiles sold. Suppose we consider the experiment of observing a day of operations at DiCarlo Motors and define the random variable of interest as x = the number of automobiles sold during a day. Using the relative frequencies to assign probabilities to the values of the random variable x , we can develop the probability distribution for x .

TABLE 5.3 PROBABILITY DISTRIBUTION FOR NUMBER OBTAINED ON ONE ROLL OF A DIE

Number Obtained	x	Probability of x	$f(x)$
1	1	1/6	1/6
2	2	1/6	1/6
3	3	1/6	1/6
4	4	1/6	1/6
5	5	1/6	1/6
6	6	1/6	1/6

TABLE 5.4 PROBABILITY DISTRIBUTION FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS

x	$f(x)$
0	.18
1	.39
2	.24
3	.14
4	.04
5	.01
Total	1.00

In probability function notation, $f(0)$ provides the probability of 0 automobiles sold, $f(1)$ provides the probability of 1 automobile sold, and so on. Because historical data show 54 of 300 days with 0 automobiles sold, we assign the relative frequency $54/300 = .18$ to $f(0)$, indicating that the probability of 0 automobiles being sold during a day is .18. Similarly, because 117 of 300 days had 1 automobile sold, we assign the relative frequency $117/300 = .39$ to $f(1)$, indicating that the probability of exactly 1 automobile being sold during a day is .39. Continuing in this way for the other values of the random variable, we compute the values for $f(2)$, $f(3)$, $f(4)$, and $f(5)$ as shown in Table 5.4.

A primary advantage of defining a random variable and its probability distribution is that once the probability distribution is known, it is relatively easy to determine the probability of a variety of events that may be of interest to a decision maker. For example, using the probability distribution for DiCarlo Motors as shown in Table 5.4, we see that the most probable number of automobiles sold during a day is 1 with a probability of $f(1) = .39$. In addition, the probability of selling 3 or more automobiles during a day is $f(3) + f(4) + f(5) = .14 + .04 + .01 = .19$. These probabilities, plus others the decision maker may ask about, provide information that can help the decision maker understand the process of selling automobiles at DiCarlo Motors.

In the development of a probability function for any discrete random variable, the following two conditions must be satisfied.

REQUIRED CONDITIONS FOR A DISCRETE PROBABILITY FUNCTION

$$f(x) \geq 0 \quad (5.1)$$

$$\sum f(x) = 1 \quad (5.2)$$

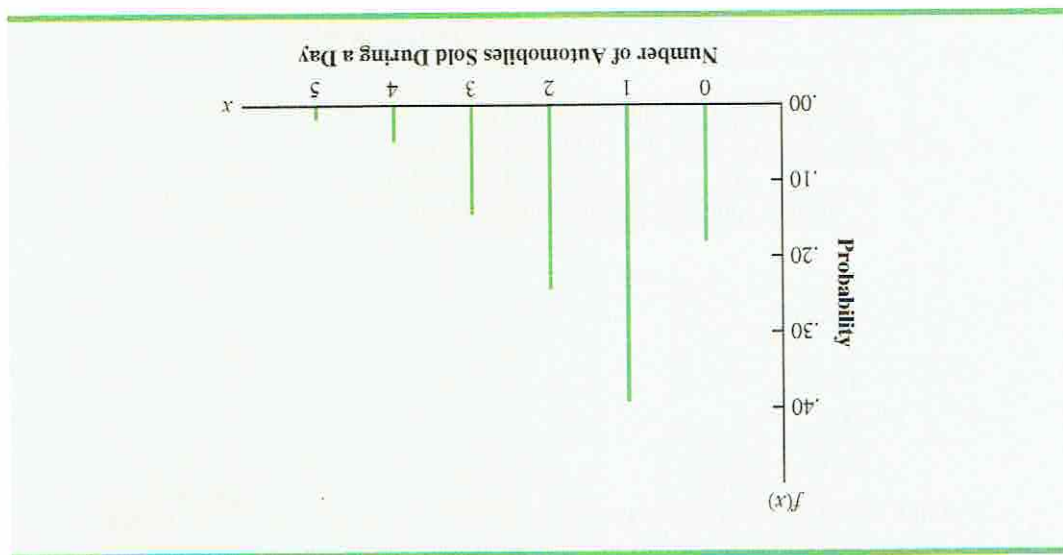
Table 5.4 shows that the probabilities for the random variable x satisfy equation (5.1); $f(x)$ is greater than or equal to 0 for all values of x . In addition, because the probabilities sum to 1, equation (5.2) is satisfied. Thus, the DiCarlo Motors probability function is a valid discrete probability function.

We can also show the DiCarlo Motors probability distribution graphically. In Figure 5.1 the values of the random variable x for DiCarlo Motors are shown on the horizontal axis and the probability associated with these values is shown on the vertical axis.

In addition to the probability distributions shown in tables, a formula that gives the probability function, $f(x)$, for every value of x is often used to describe probability

These conditions are the analogs to the two basic requirements for assigning probabilities to experimental outcomes presented in Chapter 4.

FIGURE 5.1 GRAPHICAL REPRESENTATION OF THE PROBABILITY DISTRIBUTION FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS



distributions. The simplest example of a discrete probability distribution given by a formula is the **discrete uniform probability distribution**. Its probability function is defined by equation (5.3).

DISCRETE UNIFORM PROBABILITY FUNCTION

$$f(x) = 1/n$$

(5.3)

where

n = the number of values the random variable may assume

For example, consider again the experiment of rolling a die. We define the random variable x to be the number of dots on the upward face. For this experiment, $n = 6$ values are possible for the random variable: $x = 1, 2, 3, 4, 5, 6$. We showed earlier how the probability distribution for this experiment can be expressed as a table. Since the probabilities are equally likely, the discrete uniform probability function can also be used. The probability function for this discrete uniform random variable is

$$f(x) = 1/6 \quad x = 1, 2, 3, 4, 5, 6$$

Several widely-used discrete probability distributions are specified by formulas. Three important cases are the binomial, Poisson, and hypergeometric distributions; these distributions are discussed later in the chapter.

10. The percent frequency distributions of job satisfaction scores for a sample of information systems (IS) senior executives and middle managers are as follows. The scores range from a low of 1 (very dissatisfied) to a high of 5 (very satisfied).

Job Satisfaction Score	IS Senior Executives (%)	IS Middle Managers (%)
5	41	28
4	42	46
3	3	12
2	9	10
1	5	4

- Develop a probability distribution for the job satisfaction score of a senior executive.
- Develop a probability distribution for the job satisfaction score of a middle manager.
- What is the probability a senior executive will report a job satisfaction score of 4 or 5?
- What is the probability a middle manager is very satisfied?
- Compare the overall job satisfaction of senior executives and middle managers.

11. A technician services mailing machines at companies in the Phoenix area. Depending on the type of malfunction, the service call can take 1, 2, 3, or 4 hours. The different types of malfunctions occur at about the same frequency.
 - Develop a probability distribution for the duration of a service call.
 - Draw a graph of the probability distribution.
 - Show that your probability distribution satisfies the conditions required for a discrete probability function.
 - What is the probability a service call will take 3 hours?
 - A service call has just come in, but the type of malfunction is unknown. It is 3:00 p.m. and service technicians usually get off at 5:00 p.m. What is the probability the service technician will have to work overtime to fix the machine today?
12. Time Warner Cable provides television and Internet service to millions of customers. Suppose that the management of Time Warner Cable subjectively assesses a probability distribution for the number of new subscribers next year in the state of New York as follows.

x	f(x)
100,000	.10
200,000	.20
300,000	.25
400,000	.30
500,000	.10
600,000	.05

- Is this probability distribution valid? Explain.
 - What is the probability Time Warner will obtain more than 400,000 new subscribers?
 - What is the probability Time Warner will obtain fewer than 200,000 new subscribers?
13. A psychologist determined that the number of sessions required to obtain the trust of a new patient is either 1, 2, or 3. Let x be a random variable indicating the number of sessions required to gain the patient's trust. The following probability function has been proposed.
- $$f(x) = \frac{6}{x} \quad \text{for } x = 1, 2, \text{ or } 3$$