5 Discrete Probability Distributions

CHAPTER CONTENTS

LEARNING OBJECTIVES

After reading this chapter and doing the exercises you should be able to:

1. Understand the concepts of a random variable and a probability distribution.
2. Distinguish between discrete and continuous random variables.
3. Compute and interpret the expected value, variance and standard deviation for a discrete random variable.
4. Compute and work with probabilities involving a binomial probability distribution.
5. Compute and work with probabilities involving a Poisson probability distribution.
6. Know when and how to use the hypergeometric probability distribution.

In this chapter we continue the study of probability by introducing the concepts of random variables and probability distributions. The focus of this chapter is discrete probability distributions. Three special discrete probability distributions – the binomial, Poisson and hypergeometric – are covered.

5.1 RANDOM VARIABLES

In Chapter 4 we defined the concept of an experiment and its associated experimental outcomes. A **random variable** provides a means for describing experimental outcomes using numerical values. Random variables must assume numerical values.

Random variable

A random variable is a numerical description of the outcome of an experiment.

In effect, a random variable associates a numerical value with each possible experimental outcome. The particular numerical value of the random variable depends on the outcome of the experiment. A random variable can be classified as being either *discrete* or *continuous* depending on the numerical values it assumes.

STATISTICS IN PRACTICE

Improving the performance reliability of combat aircraft

Modern combat aircraft are expensive to acquire and maintain. In today's post-Cold War world the emphasis is therefore on deploying as few aircraft as are required and for these to be made to perform as reliably as possible in conflict and peace-keeping situations. Different strategies have been considered by manufacturers for improving the performance reliability of aircraft. One such is to reduce the incidence of faults per flying hour to improve the aircraft's survival time. For example, the Tornado averages 800 faults per 1000 flying hours but if this rate could be halved, the mean operational time between faults would double. Another strategy is to build 'redundancy' into the design. In practice this would involve the aircraft carrying additional engines which would only come into use if one of the operational engines failed. To determine the number of additional engines required, designers have relied on the Poisson distribution. Calculations based on this distribution show that an aircraft with two engines would need at least four redundant engines to achieve a target maintenance-free operating period (MFOP) of 150 hours. Given that each engine weighs over a tonne, occupies a space of at least 2m^3 and costs some €3m, clearly this has enormous implications for those wishing to pursue this solution further.

Source: Kumar U D, Knezivic J and Crocker (1999) Maintenance-free operating period –an alternative measure to MTBF and failure rate for specifying reliability. Reliability Engineering & System Safety Vol 64 pp 127–131.

Discrete random variables

A random variable that may assume either a finite number of values or an infinite sequence of values such as 0, 1, 2, … is referred to as a **discrete random variable**. For example, consider the experiment of an accountant taking the chartered accountancy (CA) examination.

The examination has four parts. We can define a random variable as $X =$ the number of parts of the CA examination passed. It is a discrete random variable because it may assume the finite number of values 0, 1, 2, 3 or 4.

As another example of a discrete random variable, consider the experiment of cars arriving at a tollbooth. The random variable of interest is $X =$ the number of cars arriving during a one-day period. The possible values for X come from the sequence of integers 0, 1, 2 and so on. Hence, X is a discrete random variable assuming one of the values in this infinite sequence. Although the outcomes of many experiments can naturally be described by numerical values, others cannot. For example, a survey question might ask an individual to recall the message in a recent television commercial. This experiment would have two possible outcomes: the individual cannot recall the message and the individual can recall the message.

We can still describe these experimental outcomes numerically by defining the discrete random variable X as follows: let $X = 0$ if the individual cannot recall the message and $X = 1$ if the individual can recall the message. The numerical values for this random variable are arbitrary (we could use 5 and 10), but they are acceptable in terms of the definition of a random variable – namely, X is a random variable because it provides a numerical description of the outcome of the experiment.

Table 5.1 provides some additional examples of discrete random variables. Note that in each example the discrete random variable assumes a finite number of values or an infinite sequence of values such as 0, 1, 2, …. These types of discrete random variables are discussed in detail in this chapter.

Continuous random variables

A random variable that may assume any numerical value in an interval or collection of intervals is called a **continuous random variable**. Experimental outcomes based on measurement scales such as time, weight, distance and temperature can be described by continuous random variables. For example, consider an experiment of monitoring incoming telephone calls to the claims office of a major insurance company. Suppose the random variable of interest is $X =$ the time between consecutive incoming calls in minutes. This random variable may assume any value in the interval $X \geq 0$. Actually, an infinite number of values are possible for X, including values such as 1.26 minutes, 2.751 minutes, 4.3333 minutes and so on. As another example, consider a 90-kilometre section of the A8 Autobahn in Germany.

TABLE 5.1 Examples of discrete random variables

Experiment	Random variable (X)	Possible values for the random variable
Contact five customers	Number of customers who place an order	0, 1, 2, 3, 4, 5
Inspect a shipment of 50 radios	Number of defective radios	0, 1, 2, …, 49, 50
Operate a restaurant for one day	Number of customers	0, 1, 2, 3, …
Sell a car	Gender of the customer	0 if male; 1 if female

TABLE 5.2 Examples of continuous random variables

Experiment	Random variable (X)	Possible values for the random variable
Operate a bank	Time between customer arrivals	$X \geq 0$ *in minutes*
Fill a soft drink can (max = 350g)	Number of grams	$0 \leq X \leq 350$
Construct a new library	Percentage of project complete after six months	$0 \leq X \leq 100$
Test a new chemical process	Temperature when the desired reaction takes place (min 65°C; max 100°C)	$65 \leq X \leq 100$

For an emergency ambulance service located in Stuttgart, we might define the random variable as X = number of kilometres to the location of the next traffic accident along this section of the A8. In this case, X would be a continuous random variable assuming any value in the interval $0 \leq X \leq 90$. Additional examples of continuous random variables are listed in Table 5.2. Note that each example describes a random variable that may assume any value in an interval of values. Continuous random variables and their probability distributions will be the topic of Chapter 6.

EXERCISES

Methods

1. Consider the experiment of tossing a coin twice.

a. List the experimental outcomes.
b. Define a random variable that represents the number of heads occurring on the two tosses.
c. Show what value the random variable would assume for each of the experimental outcomes.
d. Is this random variable discrete or continuous?

2. Consider the experiment of a worker assembling a product.

a. Define a random variable that represents the time in minutes required to assemble the product.
b. What values may the random variable assume?
c. Is the random variable discrete or continuous?

COMPLETE SOLUTIONS

Applications

3. Three students have interviews scheduled for summer employment. In each case the interview results in either an offer for a position or no offer. Experimental outcomes are defined in terms of the results of the three interviews.

a. List the experimental outcomes.
b. Define a random variable that represents the number of offers made. Is the random variable continuous?

c. Show the value of the random variable for each of the experimental outcomes.

4. Suppose we know home mortgage rates for 12 Danish lending institutions. Assume that the random variable of interest is the number of lending institutions in this group that offers a 30-year fixed rate of 1.5 per cent or less. What values may this random variable assume?

COMPLETE SOLUTIONS

5. To perform a certain type of blood analysis, lab technicians must perform two procedures. The first procedure requires either one or two separate steps, and the second procedure requires either one, two or three steps.

a. List the experimental outcomes associated with performing the blood analysis.
b. If the random variable of interest is the total number of steps required to do the complete analysis (both procedures), show what value the random variable will assume for each of the experimental outcomes.

6. Listed is a series of experiments and associated random variables. In each case, identify the values that the random variable can assume and state whether the random variable is discrete or continuous.

COMPLETE SOLUTIONS

Experiment	*Random variable (X)*
a. Take a 20-question examination.	Number of questions answered correctly.
b. Observe cars arriving at a tollbooth for one hour.	Number of cars arriving at tollbooth.
c. Audit 50 tax returns.	Number of returns containing errors.
d. Observe an employee's work.	Number of non-productive hours in an eight-hour workday.
e. Weigh a shipment of goods.	Number of kilograms.

5.2 DISCRETE PROBABILITY DISTRIBUTIONS

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, the probability distribution is defined by a **probability function**, denoted by $p(x) = p(X = x)$ for all possible values, x. The probability function provides the probability for each value of the random variable. Consider the sales of cars at DiCarlo Motors in Sienna, Italy. Over the past 300 days of operation, sales data show 54 days with no cars sold, 117 days with one car sold, 72 days with two cars sold, 42 days with three cars sold, 12 days with four cars sold and three days with five cars sold. Suppose we consider the experiment of selecting a day of operation at DiCarlo Motors and define the random variable of interest as $X =$ the number of cars sold during a day. From historical data, we know X is a discrete random variable that can assume the values 0, 1, 2, 3, 4 or 5. In probability function notation, $p(0)$ provides the probability of 0 cars sold, $p(1)$ provides the probability of one car sold and so on. Because historical data show 54 of 300 days with no cars sold, we assign the value $54/300 = 0.18$ to $p(0)$, indicating that the probability of no cars being sold during a day is 0.18. Similarly, because 117 of 300 days had one car sold, we assign the value $117/300 = 0.39$ to $p(1)$, indicating that the probability of exactly one car being sold during a day is 0.39. Continuing in this way for the other values of the random variable, we compute the values for $p(2)$, $p(3)$, $p(4)$ and $p(5)$ as shown in Table 5.3, the probability distribution for the number of cars sold during a day at DiCarlo Motors.

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 5.3, we see that the most probable number of cars sold during a day is one with a probability of $p(1) = 0.39$. In addition, there is a $p(3) + p(4) + p(5) = 0.14 + 0.04 + 0.01 = 0.19$ probability of selling three or more cars during a day. These probabilities, plus others the decision-maker may ask about, provide information that can help the decision-maker understand the process of selling cars at DiCarlo Motors.

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

TABLE 5.3 Probability distribution for the number of cars sold during a day at DiCarlo Motors

x	$p(x)$
0	0.18
1	0.39
2	0.24
3	0.14
4	0.04
5	0.01
	Total 1.00

Required conditions for a discrete probability function are shown in equations (5.1) and (5.2).

$$p(x) \geq 0 \tag{5.1}$$

$$\Sigma p(x) = 1 \tag{5.2}$$

Table 5.3 shows that the probabilities for the random variable X satisfy equation (5.1); $p(x)$ 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 present probability distributions 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 tables and graphs, a formula that gives the probability function, $p(x)$, for every value of $X = x$ is often used to describe probability 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).

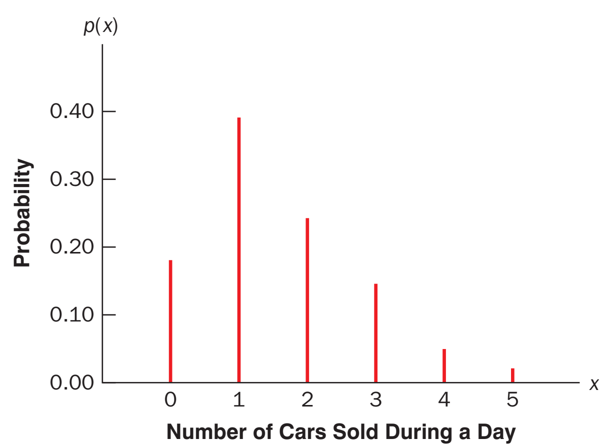
Discrete uniform probability function

$$p(x) = 1/n \tag{5.3}$$

where

n = the number of values the random variable may assume

FIGURE 5.1
Graphical representation of the probability distribution for the number of cars sold during a day at DiCarlo Motors



For example, suppose that for the experiment of rolling a die we define the random variable X to be the number of dots on the upward face. There are $n = 6$ possible values for the random variable; $X = 1, 2, 3, 4, 5, 6$. Thus, the probability function for this discrete uniform random variable is:

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

The possible values of the random variable and the associated probabilities are shown.

x	$p(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

As another example, consider the random variable X with the following discrete probability distribution.

x	$p(x)$
1	1/10
2	2/10
3	3/10
4	4/10

This probability distribution can be defined by the formula:

$$p(x) = \frac{x}{10} \quad \text{for } x = 1, 2, 3 \text{ or } 4$$

Evaluating $p(x)$ for a given value of the random variable will provide the associated probability. For example, using the preceding probability function, we see that $p(2) = 2/10$ provides the probability that the random variable assumes a value of 2. The more widely used discrete probability distributions generally are specified by formulae. Three important cases are the binomial, Poisson and hypergeometric distributions; these are discussed later in the chapter.

EXERCISES

Methods

7. The probability distribution for the random variable X follows.

x	$p(x)$
20	0.20
25	0.15
30	0.25
35	0.40

a. Is this probability distribution valid? Explain.
b. What is the probability that $X = 30$?
c. What is the probability that X is less than or equal to 25?
d. What is the probability that X is greater than 30?

Applications

8. The following data were collected by counting the number of operating rooms in use at a general hospital over a 20-day period. On three of the days only one operating room was used, on five of the days two were used, on eight of the days three were used and on four days all four of the hospital's operating rooms were used.

COMPLETE SOLUTIONS

- **a.** Use the relative frequency approach to construct a probability distribution for the number of operating rooms in use on any given day.
- **b.** Draw a graph of the probability distribution.
- **c.** Show that your probability distribution satisfies the required conditions for a valid discrete probability distribution.

9. Table 5.4 summarizes the joint probability distribution for the percentage monthly return for two ordinary shares 1 and 2. In the case of share 1, the per cent return X has historically been –1, 0 or 1. Correspondingly, for share 2, the per cent return Y has been –2, 0 or 2.

TABLE 5.4 Per cent monthly return probabilities for shares 1 and 2

			share 2	
	%		**Y**	
	Monthly return	**−2**	**0**	**2**
	−1	0.1	0.1	0.0
share 1 X	0	0.1	0.2	0.0
	1	0.0	0.1	0.4

- **a.** Determine $E(Y)$, E(X) Var(X) and Var(Y).
- **b.** Determine the correlation coefficient between X and Y.
- **c.** What do you deduce from b?

10. A technician services mailing machines at companies in the Berne 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.

COMPLETE SOLUTIONS

- **a.** Develop a probability distribution for the duration of a service call.
- **b.** Draw a graph of the probability distribution.
- **c.** Show that your probability distribution satisfies the conditions required for a discrete probability function.
- **d.** What is the probability a service call will take three hours?
- **e.** A service call has just come in, but the type of malfunction is unknown. It is 3:00 p.m. and service technicians usually finish work at 5:00 p.m. What is the probability the service technician will have to work overtime to fix the machine today?

11. A college admissions tutor subjectively assessed a probability distribution for X, the number of entering students, as follows.

x	$p(x)$
1000	0.15
1100	0.20
1200	0.30
1300	0.25
1400	0.10

a. Is this probability distribution valid? Explain.
b. What is the probability of 1200 or fewer entering students?

COMPLETE SOLUTIONS

12. 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.

$$p(x) = \frac{x}{6} \quad \text{for } x = 1, 2 \text{ or } 3$$

a. Is this probability function valid? Explain.
b. What is the probability that it takes exactly two sessions to gain the patient's trust?
c. What is the probability that it takes at least two sessions to gain the patient's trust?

13. The following table is a partial probability distribution for the MRA Company's projected profits (X = profit in €000s) for the first year of operation (the negative value denotes a loss).

x	$p(x)$
−100	0.10
0	0.20
50	0.30
100	0.25
150	0.10
200	

a. What is the proper value for $p(200)$? What is your interpretation of this value?
b. What is the probability that MRA will be profitable?
c. What is the probability that MRA will make at least €100 000?

5.3 EXPECTED VALUE AND VARIANCE

Expected value

The **expected value**, or mean, of a random variable is a measure of the central location for the random variable. The formula for the expected value of a discrete random variable X follows in equation (5.4).

Expected value of a discrete random variable

$$E(X) = \mu = \Sigma x p(x) \tag{5.4}$$

TABLE 5.5 Calculation of the expected value for the number of cars sold during a day at DiCarlo Motors

x	$p(x)$	$xp(x)$
0	0.18	$0\,(0.18) = 0.00$
1	0.39	$1\,(0.39) = 0.39$
2	0.24	$2\,(0.24) = 0.48$
3	0.14	$3\,(0.14) = 0.42$
4	0.04	$4\,(0.04) = 0.16$
5	0.01	$5(0.01) = \underline{0.05}$
		1.50
		$E(X) = \mu = \Sigma xp(x)$

Both the notations $E(\mathrm{X})$ and μ are used to denote the expected value of a random variable. Equation (5.4) shows that to compute the expected value of a discrete random variable, we must multiply each value of the random variable by the corresponding probability $p(x)$ and then add the resulting products. Using the DiCarlo Motors car sales example from Section 5.2, we show the calculation of the expected value for the number of cars sold during a day in Table 5.5. The sum of the entries in the $xp(x)$ column shows that the expected value is 1.50 cars per day. We therefore know that although sales of 0, 1, 2, 3, 4 or 5 cars are possible on any one day, over time DiCarlo can anticipate selling an average of 1.50 cars per day. Assuming 30 days of operation during a month, we can use the expected value of 1.50 to forecast average monthly sales of $30(1.50) = 45$ cars.

Variance

Even though the expected value provides the mean value for the random variable, we often need a measure of variability, or dispersion. Just as we used the variance in Chapter 3 to summarize the variability in data, we now use **variance** to summarize the variability in the values of a random variable. The formula for the variance of a discrete random variable follows in equation (5.5).

Variance of a discrete random variable

$$\mathrm{Var}(X) = \sigma^2 = \Sigma(x-\mu)^2 p(x) \tag{5.5}$$

As equation (5.5) shows, an essential part of the variance formula is the deviation, $x - \mu$, which measures how far a particular value of the random variable is from the expected value, or mean, μ. In computing the variance of a random variable, the deviations are squared and then weighted by the corresponding value of the probability function. The sum of these weighted squared deviations for all values of the random variable is referred to as the *variance*. The notations $\mathrm{Var}(X)$ and σ^2 are both used to denote the variance of a random variable.

The calculation of the variance for the probability distribution of the number of cars sold during a day at DiCarlo Motors is summarized in Table 5.6. We see that the variance is 1.25. The **standard deviation**, σ, is defined as the positive square root of the variance. Thus, the standard deviation for the number of cars sold during a day is:

$$\sigma = \sqrt{1.25} = 1.118$$

TABLE 5.6 Calculation of the variance for the number of cars sold during a day at DiCarlo Motors

x	$x - \mu$	$(x-\mu)^2$	$p(x)$	$(x-\mu)^2 p(x)$
0	$0 - 1.50 = -1.50$	2.25	0.18	$2.25 \times 0.18 = 0.4050$
1	$1 - 1.50 = -0.50$	0.25	0.39	$0.25 \times 0.39 = 0.0975$
2	$2 - 1.50 = 0.50$	0.25	0.24	$0.25 \times 0.24 = 0.0600$
3	$3 - 1.50 = 1.50$	2.25	0.14	$2.25 \times 0.14 = 0.3150$
4	$4 - 1.50 = 2.50$	6.25	0.04	$6.25 \times 0.04 = 0.2500$
5	$5 - 1.50 = 3.50$	12.25	0.01	$12.25 \times 0.01 = 0.1225$
				1.2500
				$\sigma^2 = \Sigma(x-\mu)^2 p(x)$

The standard deviation is measured in the same units as the random variable ($\sigma = 1.118$ cars) and therefore is often preferred in describing the variability of a random variable. The variance σ^2 is measured in squared units and is thus more difficult to interpret.

EXERCISES

Methods

14. The following table provides a probability distribution for the random variable X.

x	$p(x)$
3	0.25
6	0.50
9	0.25

a. Compute $E(X)$, the expected value of X.
b. Compute σ^2, the variance of X.
c. Compute σ, the standard deviation of X.

15. The following table provides a probability distribution for the random variable Y.

y	$p(y)$
2	0.20
4	0.30
7	0.40
8	0.10

a. Compute $E(Y)$.
b. Compute Var(Y) and σ.

Applications

16. Odds in horse race betting are defined as follows: 3/1 (three to one against) means a horse is expected to win once for every three times it loses; 3/2 means two wins out of five races; 4/5 (five to four on) means five wins for every four defeats, etc.

a. Translate the above odds into 'probabilities' of victory.
b. In the 2.45 race at L'Arc de Triomphe the odds for the five runners were:

Phillipe Bois	1/1
Gallante Effor	5/2
Satin Noir	11/2
Victoire Antheme	9/1
Comme Rambleur	16/1

Calculate the 'probabilities' and their sum.
c. How much would a bookmaker expect to profit in the long run at such odds if it is assumed each horse is backed equally? (Hint: Assume the true probabilities are proportional to the 'probabilities' just calculated and consider the payouts corresponding to a notional €1 wager being placed on each horse.)
d. What would the bookmaker's expected profit have been if horses had been backed in line with the true probabilities?

17. A certain machinist works an eight-hour shift. An efficiency expert wants to assess the value of this machinist where value is defined as value added minus the machinist's labour cost. The value added for the work the machinist does is €30 per item and the machinist earns €16 per hour. From past records, the machinist's output per shift is known to have the following probability distribution:

Output/shift	*Probability*
5	0.2
6	0.4
7	0.3
8	0.1

a. What is the expected monetary value of the machinist to the company per shift?
b. What is the corresponding variance value?

18. A company is contracted to finish a €100 000 project by 31 December. If it does not complete on time a penalty of €8000 per month (or part of a month) is incurred. The company estimates that if it continues alone there will be a 40 per cent chance of completing on time and that the project may be one, two, three or four months late with equal probability.

Subcontractors can be hired by the firm at a cost of €18 000. If the subcontractors are hired then the probability that the company completes on time is doubled. If the project is still late it will now be only one or two months late with equal probability.

a. Determine the expected profit when

i. subcontractors are not used
ii. subcontractors are used

b. Which is the better option for the company?

19. The following probability distributions of job satisfaction scores for a sample of information systems (IS) senior executives and IS middle managers range from a low of 1 (very dissatisfied) to a high of 5 (very satisfied).

	Probability	
Job satisfaction score	*IS senior executives*	*IS middle managers*
1	0.05	0.04
2	0.09	0.10
3	0.03	0.12
4	0.42	0.46
5	0.41	0.28

a. What is the expected value of the job satisfaction score for senior executives?
b. What is the expected value of the job satisfaction score for middle managers?
c. Compute the variance of job satisfaction scores for executives and middle managers.
d. Compute the standard deviation of job satisfaction scores for both probability distributions.
e. Compare the overall job satisfaction of senior executives and middle managers.

COMPLETE SOLUTIONS

20. The demand for a product of Cobh Industries varies greatly from month to month. The probability distribution in the following table, based on the past two years of data, shows the company's monthly demand.

Unit demand	*Probability*
300	0.20
400	0.30
500	0.35
600	0.15

a. If the company bases monthly orders on the expected value of the monthly demand, what should Cobh's monthly order quantity be for this product?
b. Assume that each unit demanded generates €70 in revenue and that each unit ordered costs €50. How much will the company gain or lose in a month if it places an order based on your answer to part (a) and the actual demand for the item is 300 units?

5.4 BINOMIAL PROBABILITY DISTRIBUTION

The binomial probability distribution is a discrete probability distribution that provides many applications. It is associated with a multiple-step experiment that we call the binomial experiment.

A binomial experiment

A **binomial experiment** exhibits the following four properties.

Properties of a binomial experiment

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes are possible on each trial. We refer to one outcome as a *success* and the other outcome as a *failure*.
3. The probability of a success, denoted by π, does not change from trial to trial. Consequently, the probability of a failure, denoted by $1 - \pi$, does not change from trial to trial.
4. The trials are independent.

FIGURE 5.2
One possible sequence of successes and failures for an eight-day trial binomial experiment

Property 1: The experiment consists of $n = 8$ identical trials.

Property 2: Each trial results in either success (*S*) or failure (*F*).

Trials →	1	2	3	4	5	6	7	8
Outcomes →	*S*	*F*	*F*	*S*	*S*	*F*	*S*	*S*

If properties 2, 3 and 4 are present, we say the trials are generated by a Bernoulli process. If, in addition, property 1 is present, we say we have a binomial experiment. Figure 5.2 depicts one possible sequence of successes and failures for a binomial experiment involving eight trials.

In a binomial experiment, our interest is in the *number of successes occurring in the* n *trials.* If we let X denote the number of successes occurring in the n trials, we see that X can assume the values of 0, 1, 2, 3, …, n. Because the number of values is finite, X is a *discrete* random variable. The probability distribution associated with this random variable is called the **binomial probability distribution**. For example, consider the experiment of tossing a coin five times and on each toss observing whether the coin lands with a head or a tail on its upward face. Suppose we want to count the number of heads appearing over the five tosses. Does this experiment show the properties of a binomial experiment? What is the random variable of interest? Note that:

1 The experiment consists of five identical trials; each trial involves the tossing of one coin.
2 Two outcomes are possible for each trial: a head or a tail. We can designate head a success and tail a failure.
3 The probability of a head and the probability of a tail are the same for each trial, with $\pi = 0.5$ and $1 - \pi = 0.5$.
4 The trials or tosses are independent because the outcome on any one trial is not affected by what happens on other trials or tosses.

Thus, the properties of a binomial experiment are satisfied. The random variable of interest is $X =$ the number of heads appearing in the five trials. In this case, X can assume the values of 0, 1, 2, 3, 4 or 5.

As another example, consider an insurance salesperson who visits ten randomly selected families. The outcome associated with each visit is classified as a success if the family purchases an insurance policy and a failure if the family does not. From past experience, the salesperson knows the probability that a randomly selected family will purchase an insurance policy is 0.10. Checking the properties of a binomial experiment, we observe that:

1 The experiment consists of ten identical trials; each trial involves contacting one family.
2 Two outcomes are possible on each trial: the family purchases a policy (success) or the family does not purchase a policy (failure).
3 The probabilities of a purchase and a non-purchase are assumed to be the same for each sales call, with $\pi = 0.10$ and $1 - \pi = 0.90$.
4 The trials are independent because the families are randomly selected.

Because the four assumptions are satisfied, this example is a binomial experiment. The random variable of interest is the number of sales obtained in contacting the ten families. In this case, X can assume the values of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

Property 3 of the binomial experiment is called the *stationarity assumption* and is sometimes confused with property 4, independence of trials. To see how they differ, consider again the case of the salesperson calling on families to sell insurance policies. If, as the day wore on, the salesperson got tired and lost enthusiasm, the probability of success (selling a policy) might drop to 0.05, for example, by the tenth call.

In such a case, property 3 (stationarity) would not be satisfied, and we would not have a binomial experiment. Even if property 4 held – that is, the purchase decisions of each family were made independently – it would not be a binomial experiment if property 3 was not satisfied.

In applications involving binomial experiments, a special mathematical formula, called the *binomial probability function,* can be used to compute the probability of x successes in the n trials. We will show in the context of an illustrative problem how the formula can be developed.

Marrine clothing store problem

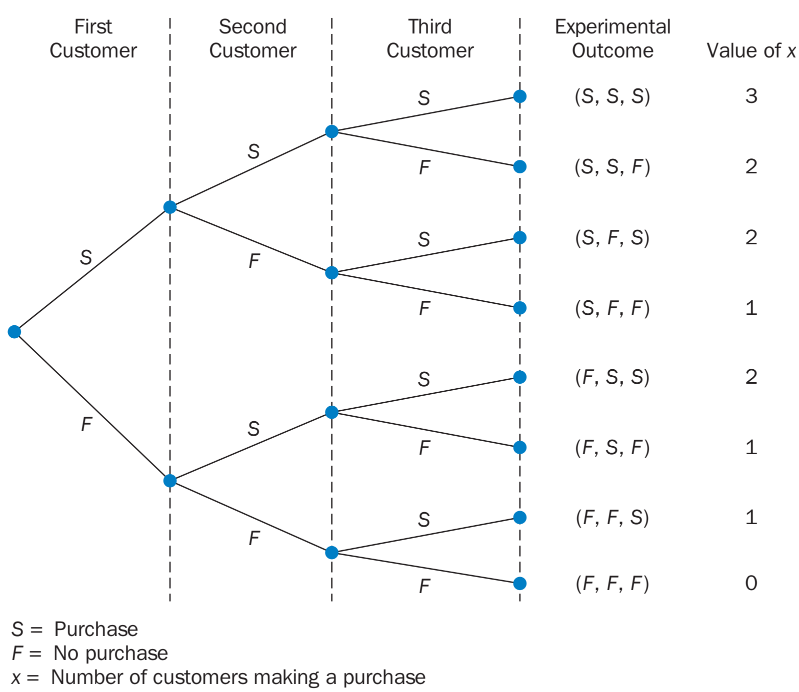
Let us consider the purchase decisions of the next three customers who enter the Marrine Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is 0.30. What is the probability that two of the next three customers will make a purchase?

Using a tree diagram (Figure 5.3), we see that the experiment of observing the three customers each making a purchase decision has eight possible outcomes. Using S to denote success (a purchase) and F to denote failure (no purchase), we are interested in experimental outcomes involving two successes in the three trials (purchase decisions). Next, let us verify that the experiment involving the sequence of three purchase decisions can be viewed as a binomial experiment. Checking the four requirements for a binomial experiment, we note that:

1. The experiment can be described as a sequence of three identical trials, one trial for each of the three customers who will enter the store.
2. Two outcomes – the customer makes a purchase (success) or the customer does not make a purchase (failure) – are possible for each trial.
3. The probability that the customer will make a purchase (0.30) or will not make a purchase (0.70) is assumed to be the same for all customers.
4. The purchase decision of each customer is independent of the decisions of the other customers.

Hence, the properties of a binomial experiment are present.

FIGURE 5.3 Tree diagram for the Marrine Clothing Store problem



The number of experimental outcomes resulting in exactly x successes in n trials can be computed using the following formula.*

Number of experimental outcomes providing exactly x successes in n trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

where

$$n! = n \times (n-1) \times (n-2) \times \ldots \times (2) \times (1)$$

and, by definition,

$$0! = 1$$

Now let us return to the Marrine Clothing Store experiment involving three customer purchase decisions. Equation (5.6) can be used to determine the number of experimental outcomes involving two purchases; that is, the number of ways of obtaining $X = 2$ successes in the $n = 3$ trials. From equation (5.6) we have:

$$\binom{n}{x} = \binom{3}{2} = \frac{3!}{2! \times (3-2)!} = \frac{3 \times 2 \times 1}{(2 \times 1) \times (1)} = \frac{6}{2} = 3$$

Equation (5.6) shows that three of the experimental outcomes yield two successes. From Figure 5.3 we see these three outcomes are denoted by (S, S, F), (S, F, S) and (F, S, S). Using equation (5.6) to determine how many experimental outcomes have three successes (purchases) in the three trials, we obtain:

$$\binom{n}{x} = \binom{3}{3} = \frac{3!}{3! \times (3-2)!} = \frac{3 \times 2 \times 1}{(3 \times 2 \times 1) \times (1)} = \frac{6}{6} = 1$$

From Figure 5.3 we see that the one experimental outcome with three successes is identified by (S, S, S).

We know that equation (5.6) can be used to determine the number of experimental outcomes that result in X successes. If we are to determine the probability of x successes in n trials, however, we must also know the probability associated with each of these experimental outcomes. Because the trials of a binomial experiment are independent, we can simply multiply the probabilities associated with each trial outcome to find the probability of a particular sequence of successes and failures.

The probability of purchases by the first two customers and no purchase by the third customer, denoted (S, S, F), is given by:

$$\pi\pi(1-\pi)$$

With a 0.30 probability of a purchase on any one trial, the probability of a purchase on the first two trials and no purchase on the third is given by:

$$0.30 \times 0.30 \times 0.70 = 0.30^2 \times 0.70 = 0.063$$

*This formula, introduced in Chapter 4, determines the number of combinations of n objects selected x at a time. For the binomial experiment, this combinatorial formula provides the number of experimental outcomes (sequences of n trials) resulting in x successes.

Two other experimental outcomes also result in two successes and one failure. The probabilities for all three experimental outcomes involving two successes follow.

Trial outcomes				
1st customer	*2nd customer*	*3rd customer*	*Experimental outcome*	*Probability of experimental outcome*
Purchase	Purchase	No purchase	(S, S, F)	$\pi\pi(1-\pi) = \pi^2(1-\pi) = (0.30)^2(0.70) = 0.063$
Purchase	No purchase	Purchase	(S, F, S)	$\pi(1-\pi)\pi = \pi^2(1-\pi) = (0.30)^2(0.70) = 0.063$
No purchase	Purchase	Purchase	(F, S, S)	$(1-\pi)\pi\pi = \pi^2(1-\pi) = (0.30)^2(0.70) = 0.063$

Observe that all three experimental outcomes with two successes have exactly the same probability. This observation holds in general. In any binomial experiment, all sequences of trial outcomes yielding x successes in n trials have the *same probability* of occurrence.

The probability of each sequence of trials yielding X successes in n trials follows in equation (5.7).

$$\text{Probability of a particular sequence of trial outcomes with } X \text{ successes in } n \text{ trials} = \pi^x(1-\pi)^{(n-x)} \qquad \textbf{(5.7)}$$

For the Marrine Clothing Store, this formula shows that any experimental outcome with two successes has a probability of $\pi^2(1-\pi)^{(3-2)} = \pi^2(1-\pi)^1 = (0.30)^2\,(0.70)^1 = 0.063$. Combining equations (5.6) and (5.7) we obtain the following **binomial probability function**.

Binomial probability function

$$p(x) = \binom{n}{x}\pi^x(1-\pi)^{(n-x)} \qquad \textbf{(5.8)}$$

where $p(x)$ = the probability of x successes in n trials

n = the number of trials

$\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$

π = the probability of a success on any one trail

$1-\pi$ = the probability of a failure on any one trail

In the Marrine Clothing Store example, we can use this function to compute the probability that no customer makes a purchase, exactly one customer makes a purchase, exactly two customers make a purchase and all three customers make a purchase. The calculations are summarized in Table 5.7, which gives the probability distribution of the number of customers making a purchase. Figure 5.4 is a graph of this probability distribution.

The binomial probability function can be applied to *any* binomial experiment. If we are satisfied that a situation demonstrates the properties of a binomial experiment and if we know the values of n and π, we can use equation (5.8) to compute the probability of x successes in the n trials.

If we consider variations of the Marrine experiment, such as ten customers rather than three entering the store, the binomial probability function given by equation (5.8) is still applicable.

TABLE 5.7 Probability distribution for the number of customers making a purchase

x	$p(x)$
0	$\frac{3!}{0!3!}(0.30)^0(0.70)^3 = 0.343$
1	$\frac{3!}{1!2!}(0.30)^1(0.70)^2 = 0.441$
2	$\frac{3!}{2!1!}(0.30)^2(0.70)^1 = 0.189$
3	$\frac{3!}{3!0!}(0.30)^3(0.70)^0 = \frac{0.027}{1.000}$

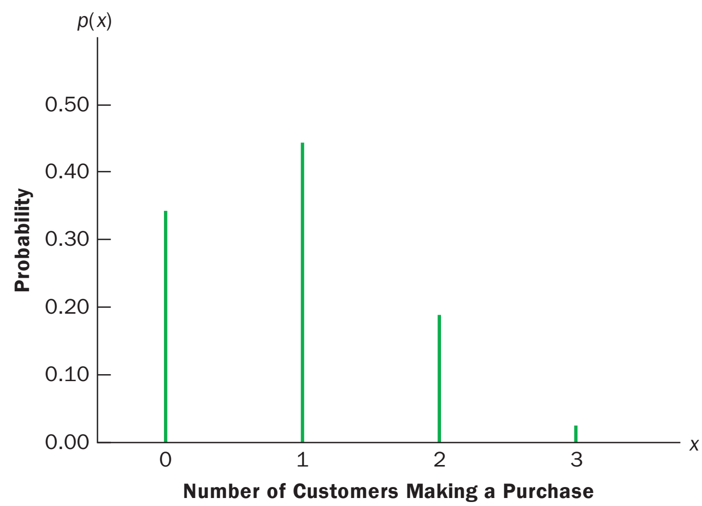


FIGURE 5.4
Graphical representation of the probability distribution for the number of customers making a purchase

Suppose we have a binomial experiment with $n = 10$, $x = 4$ and $\pi = 0.30$. The probability of making exactly four sales to ten customers entering the store is:

$$p(4) = \frac{10!}{4!6!}(0.30)^4(0.70)^6 = 0.2001$$

Using tables of binomial probabilities

Tables have been developed that give the probability of x successes in n trials for a binomial experiment. The tables are generally easy to use and quicker than equation (5.8). Table 5 of Appendix B provides such a table of binomial probabilities. To use this table, we must specify the values of n, π and x for the binomial experiment of interest. For example, the probability of $x = 3$ successes in a binomial experiment with $n = 10$ and $\pi = 0.40$ can be seen to be 0.2150. You can use equation (5.8) to verify that you would obtain the same answer if you used the binomial probability function directly.

Now let us use the same table to verify the probability of four successes in ten trials for the Marrine Clothing Store problem. Note that the value of $p(4) = 0.2001$ can be read directly from the table of binomial probabilities, with $n = 10$, $x = 4$ and $\pi = 0.30$.

Even though the tables of binomial probabilities are relatively easy to use, it is impossible to have tables that show all possible values of n and π that might be encountered in a binomial experiment. However, with today's calculators, using equation (5.8) to calculate the desired probability is not difficult, especially if the number of trials is not large. In the exercises, you should practice using equation (5.8) to compute the binomial probabilities unless the problem specifically requests that you use the binomial probability table.

FIGURE 5.5
MINITAB output showing binomial probabilities for the Marrine Clothing Store problem

x	P(X = x)
0	0.028248
1	0.121061
2	0.233474
3	0.266828
4	0.200121
5	0.102919
6	0.036757
7	0.009002
8	0.001447
9	0.000138
10	0.000006

Statistical software packages such as MINITAB, SPSS and spreadsheet packages such as EXCEL also provide a capability for computing binomial probabilities. Consider the Marrine Clothing Store example with $n = 10$ and $\pi = 0.30$. Figure 5.5 shows the binomial probabilities generated by MINITAB for all possible values of x. Note that these values are the same as those found in the $\pi = 0.30$ column of Table 5.5 of Appendix B. At the end of the chapter, details are given on how to generate the output in Figure 5.5 using first MINITAB, then EXCEL and finally SPSS.

Expected value and variance for the binomial distribution

In Section 5.3 we provided formulae for computing the expected value and variance of a discrete random variable. In the special case where the random variable has a binomial distribution with a known number of trials n and a known probability of success π, the general formulae for the expected value and variance can be simplified. The results follow.

Expected value and variance for the binomial distribution

$$E(X) = \mu = n\pi \tag{5.9}$$

$$\text{Var}(X) = \sigma^2 = n\pi(1 - \pi) \tag{5.10}$$

For the Marrine Clothing Store problem with three customers, we can use equation (5.9) to compute the expected number of customers who will make a purchase.

$$E(X) = n\pi = 3 \times 0.30 = 0.9$$

Suppose that for the next month the Marrine Clothing Store forecasts 1000 customers will enter the store. What is the expected number of customers who will make a purchase? The answer is $\mu = n\pi = 1000 \times 0.3 = 300$. Thus, to increase the expected number of purchases, Marrine must induce more customers to enter the store and/or somehow increase the probability that any individual customer will make a purchase after entering.

For the Marrine Clothing Store problem with three customers, we see that the variance and standard deviation for the number of customers who will make a purchase are:

$$\sigma^2 = n\pi(1 - \pi) = 3 \times 0.3 \times 0.7 = 0.63$$

$$\sigma = \sqrt{0.063} = 0.79$$

For the next 1000 customers entering the store, the variance and standard deviation for the number of customers who will make a purchase are:

$$\sigma^2 = n\pi(1 - \pi) = 1000 \times 0.3 \times 0.7 = 210$$

$$\sigma = \sqrt{210} = 14.49$$

EXERCISES

Methods

21. Consider a binomial experiment with two trials and $\pi = 0.4$.

a. Draw a tree diagram for this experiment (see Figure 5.3).
b. Compute the probability of one success, $p(1)$.
c. Compute $p(0)$.
d. Compute $p(2)$.
e. Compute the probability of at least one success.
f. Compute the expected value, variance and standard deviation.

22. Consider a binomial experiment with $n = 10$ and $\pi = 0.10$.

a. Compute $p(0)$.
b. Compute $p(2)$.
c. Compute $P(x \leq 2)$.
d. Compute $P(x \geq 1)$.
e. Compute $E(X)$.
f. Compute Var(X) and σ.

23. Consider a binomial experiment with $n = 20$ and $\pi = 0.70$.

a. Compute $p(12)$.
b. Compute $p(16)$.
c. Compute $P(X \geq 16)$.
d. Compute $P(X \leq 15)$.
e. Compute $E(X)$.
f. Compute Var(X) and σ.

Applications

24. When a new machine is functioning properly, only 3 per cent of the items produced are defective. Assume that we will randomly select two parts produced on the machine and that we are interested in the number of defective parts found.

a. Describe the conditions under which this situation would be a binomial experiment.
b. Draw a tree diagram similar to Figure 5.3 showing this problem as a two-trial experiment.
c. How many experimental outcomes result in exactly one defect being found?
d. Compute the probabilities associated with finding no defects, exactly one defect and two defects.

25. It takes at least nine votes from a 12-member jury to convict a defendant. Suppose that the probability that a juror votes a guilty person innocent is 0.2 whereas the probability that the juror votes an innocent person guilty is 0.1.

a. If each juror acts independently and 65 per cent of defendants are guilty, what is the probability that the jury renders a correct decision.
b. What percentage of defendants is convicted?

26. A firm bills its accounts at a 1 per cent discount for payment within ten days and the full amount is due after ten days. In the past 30 per cent of all invoices have been paid within ten days. If the firm sends out eight invoices during the first week of January, what is the probability that:

a. No one receives the discount?
b. Everyone receives the discount?
c. No more than three receive the discount?
d. At least two receive the discount?

27. In a game of 'Chuck a luck' a player bets on one of the numbers 1 to 6. Three dice are then rolled and if the number bet by the player appears *i* times ($i = 1, 2, 3$) the player then wins *i* units. On the other hand if the number bet by the player does not appear on any of the dice the player loses 1 unit. If *x* is the player's winnings in the game, what is the expected value of *X*?

5.5 POISSON PROBABILITY DISTRIBUTION

In this section we consider a discrete random variable that is often useful in estimating the number of occurrences over a specified interval of time or space. For example, the random variable of interest might be the number of arrivals at a car wash in one hour, the number of repairs needed in ten kilometres of highway, or the number of leaks in 100 kilometres of pipeline.

If the following two properties are satisfied, the number of occurrences is a random variable described by the **Poisson probability distribution**.

Properties of a Poisson experiment

1. The probability of an occurrence is the same for any two intervals of equal length.
2. The occurrence or non-occurrence in any interval is independent of the occurrence or non-occurrence in any other interval.

The **Poisson probability function** is defined by equation (5.11).

Poisson probability function

$$p(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{5.11}$$

where

$p(x)$ = the probability of x occurrences in an interval

μ = expected value or mean number of occurrences in an interval

e = 2.71828

Before we consider a specific example to see how the Poisson distribution can be applied, note that the number of occurrences, x, has no upper limit. It is a discrete random variable that may assume an infinite sequence of values ($x = 0, 1, 2, \ldots$).

An example involving time intervals

Suppose that we are interested in the number of arrivals at the payment kiosk of a car park during a 15-minute period on weekday mornings. If we can assume that the probability of a car arriving is the same for any two time periods of equal length and that the arrival or non-arrival of a car in any time period is independent of the arrival or non-arrival in any other time period, the Poisson probability function is applicable. Suppose these assumptions are satisfied and an analysis of historical data shows that the average number of cars arriving in a 15-minute period of time is ten; in this case, the following probability function applies:

$$p(x) = \frac{10^x e^{-10}}{x!}$$

The random variable here is $X =$ number of cars arriving in any 15-minute period.

If management wanted to know the probability of exactly five arrivals in 15 minutes, we would set $X = 5$ and thus obtain:

$$\text{Probability of exactly five arrivals in 15 minutes} = p(5) = \frac{10^5 e^{-10}}{5!} = 0.0378$$

Although this probability was determined by evaluating the probability function with $\mu = 10$ and $x = 5$, it is often easier to refer to a table for the Poisson distribution. The table provides probabilities for specific values of x and μ. We include such a table as Table 7 of Appendix B. Note that to use the table of Poisson probabilities, we need know only the values of x and μ. From this table we see that the probability of five arrivals in a 15-minute period is found by locating the value in the row of the table corresponding to $x = 5$ and the column of the table corresponding to $\mu = 10$. Hence, we obtain $p(5) = 0.0378$.

In the preceding example, the mean of the Poisson distribution is $\mu = 10$ arrivals per 15-minute period. A property of the Poisson distribution is that the mean of the distribution and the variance of the distribution are *equal*. Thus, the variance for the number of arrivals during 15-minute periods is $\sigma^2 = 10$. The standard deviation is:

$$\sigma = \sqrt{10} = 3.16$$

Our illustration involves a 15-minute period, but other time periods can be used. Suppose we want to compute the probability of one arrival in a three-minute period. Because ten is the expected number of arrivals in a 15-minute period, we see that $10/15 = 2/3$ is the expected number of arrivals in a one-minute period and that $2/3 \times 3$ minutes $= 2$ is the expected number of arrivals in a three-minute period. Thus, the probability of x arrivals in a three-minute time period with $\mu = 2$ is given by the following Poisson probability function.

$$p(x) = \frac{2^x e^{-2}}{x!}$$

The probability of one arrival in a three-minute period is calculated as follows:

$$\text{Probability of exactly one arrival in three minutes} = P(1) = \frac{2^1 e^{-2}}{1!} = 0.2707$$

Earlier we computed the probability of five arrivals in a 15-minute period; it was 0.0378. Note that the probability of one arrival in a three-minute period (0.2707) is not the same. When computing a Poisson probability for a different time interval, we must first convert the mean arrival rate to the time period of interest and then compute the probability.

An example involving length or distance intervals

Consider an application not involving time intervals in which the Poisson distribution is useful. Suppose we are concerned with the occurrence of major defects in a highway, one month after resurfacing. We will assume that the probability of a defect is the same for any two highway intervals of equal length and that the occurrence or non-occurrence of a defect in any one interval is independent of the occurrence or non-occurrence of a defect in any other interval. Hence, the Poisson distribution can be applied.

Suppose that major defects one month after resurfacing occur at the average rate of two per kilometre. Let us find the probability of no major defects in a particular three-kilometre section of the highway. Because we are interested in an interval with a length of three kilometres, $\mu = 2$ defects/kilometre $\times$ 3 kilometres $= 6$ represents the expected number of major defects over the three-kilometre section of highway. Using equation (5.11), the probability of no major defects is $p(0) = 6^0 e^{-6}/0! = 0.0025$. Thus, it is unlikely that no major defects will occur in the three-kilometre section. Equivalently there is a $1 - 0.0025 = 0.9975$ probability of at least one major defect in the three-kilometre highway section.

EXERCISES

Methods

COMPLETE SOLUTIONS

28. Consider a Poisson distribution with $\mu = 3$.

a. Write the appropriate Poisson probability function.
b. Compute $p(2)$.
c. Compute $p(1)$.
d. Compute $P(X \geq 2)$.

29. Consider a Poisson distribution with a mean of two occurrences per time period.

a. Write the appropriate Poisson probability function.
b. What is the expected number of occurrences in three time periods?
c. Write the appropriate Poisson probability function to determine the probability of x occurrences in three time periods.
d. Compute the probability of two occurrences in one time period.
e. Compute the probability of six occurrences in three time periods.
f. Compute the probability of five occurrences in two time periods.

Applications

30. A certain process produces 100m long rolls of high quality silk. In order to assess quality a 10m sample is taken from the end of each roll and inspected for blemishes. The number of blemishes in each sample is thought to follow a Poisson distribution with an average of two blemishes per 10m sample.

a. What is the probability that there will be more than seven blemishes if a 30m sample is taken?

31. During the period of time that a local university takes phone-in registrations, calls come in at the rate of one every two minutes.

a. What is the expected number of calls in one hour?
b. What is the probability of three calls in five minutes?
c. What is the probability of no calls in a five-minute period?

32. Airline passengers arrive randomly and independently at the passenger-screening facility at a major international airport. The mean arrival rate is ten passengers per minute.

a. Compute the probability of no arrivals in a one-minute period.
b. Compute the probability that three or fewer passengers arrive in a one-minute period.
c. Compute the probability of no arrivals in a 15-second period.
d. Compute the probability of at least one arrival in a 15-second period.

5.6 HYPERGEOMETRIC PROBABILITY DISTRIBUTION

The **hypergeometric probability distribution** is closely related to the binomial distribution. The two probability distributions differ in two key ways. With the hypergeometric distribution, the trials are not independent; and the probability of success changes from trial to trial.

In the usual notation for the hypergeometric distribution, r denotes the number of elements in the population of size N labelled success, and $N - r$ denotes the number of elements in the population labelled failure. The **hypergeometric probability function** is used to compute the probability that in a random selection of n elements, selected without replacement, we obtain x elements labelled success and $n - x$ elements labelled failure. For this outcome to occur, we must obtain x successes from the r

successes in the population and $n - x$ failures from the $N - r$ failures. The following hypergeometric probability function provides $p(x)$, the probability of obtaining x successes in a sample of size n.

Hypergeometric probability function

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \tag{5.12}$$

where

$p(x)$ = probability of x successes in n trials
n = number of trials
N = number of elements in the population
r = number of elements in the population labelled success

Note that $\binom{N}{n}$ represents the number of ways a sample of size n can be selected from a population of size N; $\binom{r}{x}$ represents the number of ways that x successes can be selected from a total of r successes in the population; and $\binom{N-r}{n-x}$ represents the number of ways that $n - x$ failures can be selected from a total of $N - r$ failures in the population.

To illustrate the computations involved in using equation (5.12), consider the following quality control application. Electric fuses produced by Warsaw Electric are packaged in boxes of 12 units each. Suppose an inspector randomly selects three of the 12 fuses in a box for testing. If the box contains exactly five defective fuses, what is the probability that the inspector will find exactly one of the three fuses defective? In this application, $n = 3$ and $N = 12$. With $r = 5$ defective fuses in the box the probability of finding $x = 1$ defective fuse is:

$$p(1) = \frac{\binom{5}{1}\binom{7}{2}}{\binom{12}{3}} = \frac{\frac{5!}{1!4!}\,\frac{7!}{2!3!}}{\frac{12!}{3!9!}} = \frac{5 \times 21}{220} = 0.4733$$

Now suppose that we wanted to know the probability of finding *at least* one defective fuse. The easiest way to answer this question is to first compute the probability that the inspector does not find any defective fuses. The probability of $x = 0$ is:

$$p(0) = \frac{\binom{5}{0}\binom{7}{3}}{\binom{12}{3}} = \frac{\frac{5!}{0!5!}\,\frac{7!}{3!4!}}{\frac{12!}{3!9!}} = \frac{1 \times 35}{220} = 0.1591$$

With a probability of zero defective fuses $p(0) = 0.1591$, we conclude that the probability of finding at least one defective fuse must be $1 - 0.1591 = 0.8409$. Thus, there is a reasonably high probability that the inspector will find at least one defective fuse.

The mean and variance of a hypergeometric distribution are as follows.

Expected value for the hypergeometric distribution

$$E(x) = \mu = n\left(\frac{r}{N}\right) \tag{5.13}$$

Variance for the hypergeometric distribution

$$\text{Var}(X) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{5.14}$$

In the preceding example $n = 3$, $r = 5$, and $N = 12$. Thus, the mean and variance for the number of defective fuses is:

$$\mu = n\left(\frac{r}{N}\right) = 3\left(\frac{5}{12}\right) = 1.25$$

$$\sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) = 3\left(\frac{5}{12}\right)\left(1 - \frac{5}{12}\right)\left(\frac{12-3}{12-1}\right) = 0.60$$

The standard deviation is:

$$\sigma = \sqrt{0.60} = 0.77$$

EXERCISES

Methods

33. Suppose $N = 10$ and $r = 3$. Compute the hypergeometric probabilities for the following values of n and x.

a. $n = 4$, $x = 1$.
b. $n = 2$, $x = 2$.
c. $n = 2$, $x = 0$.
d. $n = 4$, $x = 2$.

34. Suppose $N = 15$ and $r = 4$. What is the probability of $x = 3$ for $n = 10$?

COMPLETE SOLUTIONS

Applications

35. Blackjack, or Twenty-one as it is frequently called, is a popular gambling game played in Monte Carlo casinos. A player is dealt two cards. Face cards (jacks, queens and kings) and tens have a point value of ten. Aces have a point value of one or 11. A 52-card deck contains 16 cards with a point value of ten (jacks, queens, kings and tens) and four aces.

a. What is the probability that both cards dealt are aces or ten-point cards?
b. What is the probability that both of the cards are aces?
c. What is the probability that both of the cards have a point value of ten?
d. A blackjack is a ten-point card and an ace for a value of 21. Use your answers to parts (a), (b) and (c) to determine the probability that a player is dealt a blackjack. (Hint: Part (d) is not a hypergeometric problem. Develop your own logical relationship as to how the hypergeometric probabilities from parts (a), (b) and (c) can be combined to answer this question.)

36. A company plans to select a team of five students from Gulf University for a business game competition from a pool of 18 undergraduates. Nine are from the second-year management course, five are third-year management and the remainder are from outside the management school. What is the probability that:

a. All five team members are second-year management?
b. No students from outside the management school are selected?

37. Manufactured parts are shipped in lots of 15 items. Four parts are randomly drawn from each lot and tested and the lot is considered acceptable if no defectives are among the four tested.

a. What is the probability that the shipment will be rejected?

ONLINE RESOURCES

For the data files, additional online summary, questions, answers and the software section for this chapter, go to the online platform.

SUMMARY

A random variable provides a numerical description of the outcome of an experiment. The probability distribution for a random variable describes how the probabilities are distributed over the values the random variable can assume. A variety of examples are used to distinguish between discrete and continuous random variables. For any discrete random variable X, the probability distribution is defined by a probability function, denoted by $p(x) = p(X = x)$, which provides the probability associated with each value of the random variable. From the probability function, the expected value, variance and standard deviation for the random variable can be computed and relevant interpretations of these terms are provided.

Particular attention was devoted to the binomial distribution which can be used to determine the probability of x successes in n trials whenever the experiment has the following properties:

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes are possible on each trial, one called success and the other failure.
3. The probability of a success π does not change from trial to trial. Consequently, the probability of failure, $1 - \pi$, does not change from trial to trial.
4. The trials are independent.

Formulae were also presented for the probability function, mean and variance of the binomial distribution.

The Poisson distribution can be used to determine the probability of obtaining x occurrences over an interval of time or space. The necessary assumptions for the Poisson distribution to apply in a given situation are that:

1. The probability of an occurrence of the event is the same for any two intervals of equal length.
2. The occurrence or non-occurrence of the event in any interval is independent of the occurrence or non-occurrence of the event in any other interval.

A third discrete probability distribution, the hypergeometric, was introduced in Section 5.6. Like the binomial, it is used to compute the probability of x successes in n trials. But, in contrast to the binomial, the probability of success changes from trial to trial.

KEY TERMS

Binomial experiment
Binomial probability distribution
Binomial probability function
Continuous random variable
Discrete random variable
Discrete uniform probability distribution
Expected value
Hypergeometric probability distribution
Hypergeometric probability function
Poisson probability distribution
Poisson probability function
Probability distribution
Probability function
Random variable
Standard deviation
Variance

KEY FORMULAE

Discrete uniform probability function

$$p(x) = 1/n \tag{5.3}$$

where

$$n = \text{the number of values the random variable may assume}$$

Expected value of a discrete random variable

$$E(X) = \mu = \Sigma x p(x) \tag{5.4}$$

Variance of a discrete random variable

$$\text{Var}(X) = \sigma^2 = \Sigma(x-\mu)^2 p(x) \tag{5.5}$$

Number of experimental outcomes providing exactly x successes in n trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

Binomial probability function

$$p(x) = \binom{n}{x}\pi^2(1-\pi)^{(n-x)} \tag{5.8}$$

Expected value for the binomial distribution

$$E(X) = \mu = n\pi \tag{5.9}$$

Variance for the binomial distribution

$$\mathrm{Var}(X)\sigma^2 = n\pi(1-\pi) \tag{5.10}$$

Poisson probability function

$$p(x) = \frac{\mu^x e^{-\mu}}{\mathrm{x}!} \tag{5.11}$$

Hypergeometric probability function

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \tag{5.12}$$

Expected value for the hypergeometric distribution

$$E(x) = \mu = n\left(\frac{r}{N}\right) \tag{5.13}$$

Variance for the hypergeometric distribution

$$Var(X) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{5.14}$$

CASE PROBLEM 1

Adapting a Bingo Game

Gaming Machines International (GMI) is investigating the adaptation of one of its bingo machine formats to allow for a bonus game facility. With the existing setup, the player has to select seven numbers from the series 1 to 80. Fifteen numbers are then drawn randomly from the 80 available and prizes awarded, according to how many of the 15 coincide with the player's selection, as follows:

Number of 'hits'	*Payoff*
0	0
1	0
2	0
3	1
4	10
5	100
6	1 000
7	100 000

With the new 'two ball bonus draw' feature, players effectively have the opportunity to improve their prize by buying an extra two balls. Note, however, that the bonus draw is only expected to be available to players who have scored 2, 3, 4 or 5 hits in the main game.

Managerial report

1. Determine the probability characteristics of GMI's original bingo game and calculate the player's expected payoff.
2. Derive corresponding probability details for the proposed bonus game. What is the probability of the player scoring:

   a. 0 hits

   b. 1 hit

   c. 2 hits

   with the extra two balls?
3. Use the results obtained from two to revise the probability distribution found for one. Hence calculate the player's expected payoff in the enhanced game. Comment on how much the player might be charged for the extra gamble.

CASE PROBLEM 2

European Airline Overbooking

EU Regulation 261/2004 sets the minimum levels of passenger compensation for denied boarding due to overbooking, and extends its coverage to include flight cancellations and long delays. It is estimated that the annual cost to airlines over and above existing compensation will total €560 million for all EU airlines:

- Compensation for overbooking affects around 250 000 passengers (0.1 per cent of total). Higher compensation rates will add €96 million to airline costs.
- At an estimated €283 million and €176 million respectively, compensation for long delays and cancellation threaten to add most additional costs incurred by European airlines. The cost to a medium-sized European airline has been estimated at €40 million a year. That represents around 20 per cent of 2004 operating profits.

EA is a small, short-range airline headquartered in Vienna. It has a fleet of small Fokker planes with a capacity of 80 passengers each. They do not have different classes in their planes. In planning for their financial obligations, EA has requested a study of the chances of 'bumping' passengers they have to consider for their overbooking strategy. The airline reports a historical 'no shows' history of 10 per cent to 12 per cent. Compensation has been set at €250 per passenger denied boarding.

Managerial report

Write a report giving the airline some scenarios of their options. Consider scenarios according to their

policy of the number of bookings/plane: 80, 85, 89, etc.

1. What percentage of the time should they estimate that their passengers will find a seat when they show up?
2. What percentage of the time some passengers may not find a seat?
3. In each case you consider, find the average amount of loss per plane they have to take into account.