Some important discrete probability distributions

CHAPTER 5

Advertising by Gaia Cruises, introduced in Chapter 4, claims that you can plan and book a cruise online anywhere and anytime, with any question being answered by a knowledgeable Australian-based sales representative within 2 hours.

Yang, who is in charge of Gaia Cruises' online enquiry and booking procedures, is investigating several key performance indicators (KPIs), in particular:

- the proportion of online enquiries that are converted to confirmed bookings
- the number of enquiries received in an hour
- the proportion of online enquiries answered within 2 hours of receipt.

Recent data collected by Yang shows that:

- 10% of enquiries are converted to confirmed bookings
- on average, Gaia Cruises receives 75 online enquiries an hour
- with the current enquiry staff levels, when
  - 24 or more enquiries are received in 12 minutes, enquiries start to queue and may not be answered within the stated 2 hours
  - fewer than five enquiries are received in 8 minutes, enquiry staff have significant idle time.

Yang would like to determine the probability of a given number of enquiries being converted to confirmed bookings in a sample of a specific size. In addition, to help determine optimal enquiry staffing levels, Yang would like to calculate the probability of receiving 24 or more enquiries in any 12 minutes and fewer than five enquiries in any 8 minutes.

Answers to these questions and others can help Gaia Cruises to develop future sales, marketing and staffing strategies.

learning objectives

After studying this chapter you should be able to:

1. recognise and use the properties of a probability distribution
2. calculate the expected value and variance of a probability distribution
3. calculate average return and measure risk associated with various investment proposals
4. identify situations that can be modelled by a binomial distribution and calculate binomial probabilities
5. identify situations that can be modelled by a Poisson distribution and calculate Poisson probabilities
6. identify situations that can be modelled by a hypergeometric distribution and calculate hypergeometric probabilities

To help answer the given probability questions Yang can use a model, or small-scale representation, that approximates the online enquiry and booking processes, allowing inferences to be made about the processes. Although model building is a difficult task for some endeavours, in this case Yang can use *probability distributions*, which are mathematical models suitable for solving these types of probability questions.

This chapter introduces probability distributions and specifically explains how to apply the binomial, Poisson and hypergeometric distributions to business and other problems.

LEARNING OBJECTIVE 1
Recognise and use the properties of a probability distribution

5.1 PROBABILITY DISTRIBUTION FOR A DISCRETE RANDOM VARIABLE

A *numerical variable* (see Chapter 1) is a variable that yields numerical responses such as the number of magazines you subscribe to or your height in centimetres. Numerical variables are classified as either *continuous* or *discrete*. Continuous numerical variables have outcomes that arise from a measuring process, for example your height or weight. Discrete numerical variables have outcomes that arise from a counting process, such as the number of magazines you subscribe to or the number of phone calls received in an hour. This chapter introduces probability distributions that represent discrete numerical variables; continuous probability distributions are discussed in Chapter 6.

probability distribution for a discrete random variable
Values of a discrete random variable with the corresponding probability of occurrence.

A **probability distribution for a discrete random variable** is a mutually exclusive list of all possible numerical outcomes of the random variable with the probability of occurrence associated with each outcome.

As an example, Table 5.1 gives the distribution of the number of home mortgages approved per week by the loans manager at a local branch of Check$mart Bank. From this we can see that the loans manager approves no more than six home mortgages per week as the list in Table 5.1 is collectively exhaustive. Furthermore, as one of the outcomes must happen – that is, between none and six mortgages approved – the probabilities must sum to 1. Figure 5.1 is a graphical representation of Table 5.1.

Home mortgages approved per week	Probability
0	0.10
1	0.10
2	0.20
3	0.30
4	0.15
5	0.10
6	0.05

Table 5.1
Probability distribution of the number of home mortgages approved per week

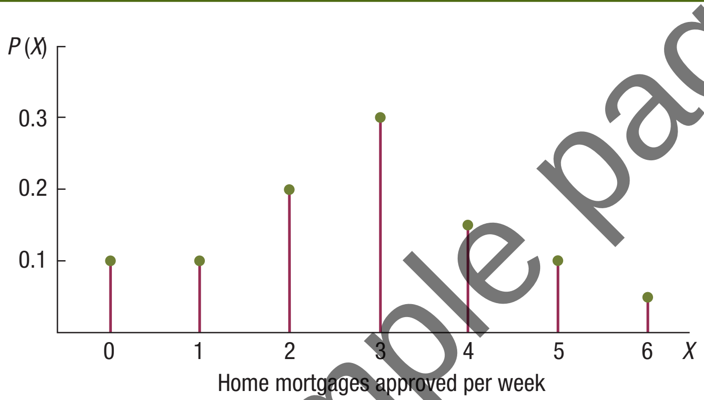

Figure 5.1
Probability distribution of the number of home mortgages approved per week

Expected Value of a Discrete Random Variable

LEARNING OBJECTIVE 2
Calculate the expected value and variance of a probability distribution

In Chapter 3 we used the sample mean and variance to describe the centre and dispersion of a sample. In the same way, we can use the mean and variance of a random variable to describe the centre and dispersion of a probability distribution.

The mean μ of a probability distribution is the *expected value* of its random variable. To calculate the **expected value of a discrete random variable** multiply each outcome X by its corresponding probability $P(X)$ and then sum these products.

expected value of a discrete random variable
Measure of central tendency; the mean of a discrete random variable.

EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i) \tag{5.1}$$

where X_i = the ith outcome of the discrete random variable X
$P(X_i)$ = probability of occurrence of the ith outcome of X

Using Equation 5.1 the mean, or expected value, for the probability distribution of the number of home mortgages approved per week is:

$$\begin{aligned}\mu &= E(X)\\ &= \sum_{i=1}^{N} X_i P(X_i)\\ &= (0 \times 0.1) + (1 \times 0.1) + (2 \times 0.2) + (3 \times 0.3) + (4 \times 0.15) + (5 \times 0.1) + (6 \times 0.05)\\ &= 0 + 0.1 + 0.4 + 0.9 + 0.6 + 0.5 + 0.3\\ &= 2.8\end{aligned}$$

The actual number of mortgages approved in a given week must be an integer value, so 2.8 mortgages are never approved in one week. However, on average, or in the long run, 2.8 are approved per week.

Variance and Standard Deviation of a Discrete Random Variable

variance of a discrete random variable
Measure of variation, based on squared deviations from the mean; directly related to the standard deviation.

standard deviation of a discrete random variable
Measure of variation, based on squared deviations from the mean; directly related to the variance.

The variance of a discrete probability distribution is calculated by multiplying each squared deviation from the mean $[X_i - E(X)]^2$ by its corresponding probability $P(X_i)$ and then summing the resulting products. Equations 5.2a and 5.3 define, respectively, the **variance of a discrete random variable** and the **standard deviation of a discrete random variable**.

VARIANCE OF A DISCRETE RANDOM VARIABLE – DEFINITION FORMULA

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i) \tag{5.2a}$$

where X_i = the ith outcome of the discrete random variable X
$P(X_i)$ = probability of occurrence of the ith outcome of X

As for sample variance, we can use algebra to obtain an alternative *calculation formula*.

VARIANCE OF A DISCRETE RANDOM VARIABLE – CALCULATION FORMULA

$$\sigma^2 = \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2 \tag{5.2b}$$

where $\sum_{i=1}^{N} X_i^2 P(X_i) = X_1^2 P(X_1) + X_2^2 P(X_2) + \ldots + X_N^2 P(X_N)$

STANDARD DEVIATION OF A DISCRETE RANDOM VARIABLE
The standard deviation of a discrete random variable is the square root of the variance

$$\sigma = \sqrt{\sigma^2} \tag{5.3}$$

Using Equations 5.2b and 5.3, the variance and standard deviation for the probability distribution of the number of mortgages approved per week are:

$$\sigma^2 = \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2$$
$$= [(0^2 \times 0.1) + (1^2 \times 0.1) + (2^2 \times 0.2) + (3^2 \times 0.3) + (4^2 \times 0.15)$$
$$+ (5^2 \times 0.1) + (6^2 \times 0.05)] - 2.8^2$$
$$= [(0 \times 0.1) + (1 \times 0.1) + (4 \times 0.2) + (9 \times 0.3) + (16 \times 0.15)$$
$$+ (25 \times 0.1) + (36 \times 0.05)] - 7.84$$
$$= (0 + 0.1 + 0.8 + 2.7 + 2.4 + 2.5 + 1.8) - 7.84$$
$$= 10.3 - 7.84$$
$$= 2.46$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{2.46} = 1.568\ldots$$

Alternatively, a table format can be used to calculate the mean and variance. In Table 5.2, the mean number of home mortgages approved per week is calculated. Then, using Equation 5.2b:

$$\sigma^2 = \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2 = 10.3 - (2.8)^2 = 2.46$$

Home mortgages approved per week X_i	$P(X_i)$	$X_i P(X_i)$	$X_i^2 P(X_i)$
0	0.10	0.0	0.0
1	0.10	0.1	0.1
2	0.20	0.4	0.8
3	0.30	0.9	2.7
4	0.15	0.6	2.4
5	0.10	0.5	2.5
6	0.05	0.3	1.8
	1.00	$\mu = E(X) = 2.8$	10.3

Table 5.2 Calculating the mean and variance of the number of home mortgages approved per week

The expected value is often used to measure the amount we can expect to gain or lose by undertaking a particular investment, while the standard deviation is used to measure the risk involved.

Problems for Section 5.1

LEARNING THE BASICS

5.1 Given the following probability distributions:

Distribution A X	$P(X)$	Distribution B X	$P(X)$
0	0.50	0	0.05
1	0.20	1	0.10
2	0.15	2	0.15
3	0.10	3	0.20
4	0.05	4	0.50

a. Calculate the expected value for each distribution.
b. Calculate the standard deviation for each distribution.
c. Compare and contrast the results of distributions *A* and *B*.

5.2 Are each of the following a valid probability distribution? Justify your answers:

Distribution A		Distribution B		Distribution C		Distribution D	
X	$P(X)$	X	$P(X)$	X	$P(X)$	X	$P(X)$
−1	0.2	0	0.1	0.250	0.500	0	0.2
1	0.9	1	0.2	0.500	0.250	1	0.1
2	−0.1	2	0.3	1.000	0.250	2	0.4
		3	0.3			3	0.5

APPLYING THE CONCEPTS

5.3 Using the company records for the past 500 working days, the manager of Konig Motors has summarised the number of cars sold per day in the following table:

Number of cars sold per day	Frequency of occurrence
0	40
1	100
2	142
3	66
4	36
5	30
6	26
7	20
8	16
9	14
10	8
11	2
Total	500

a. Form the probability distribution for the number of cars sold per day.
b. Calculate the mean or expected number of cars sold per day.
c. Calculate the standard deviation.

5.4 The manager of a large computer network has developed the following probability distribution of the number of interruptions per day:

Interruptions (X)	$P(X)$
0	0.32
1	0.35
2	0.18
3	0.08
4	0.04
5	0.02
6	0.01

a. Calculate the mean or expected number of interruptions per day.
b. Calculate the standard deviation.

5.5 In the casino version of the traditional Australian game of two-up, a spinner stands in a ring and tosses two coins into the air. The coins may land showing two heads, two tails or one tail and one head (odds). Players can bet on either heads or tails at odds of one to one. Therefore, if a player bets $1 on heads, the player will win $1 if the coins land on heads but lose $1 if the coins land on tails. Alternatively, if a player bets $1 on tails, the player will win $1 if the coins land on tails but lose $1 if the coins land on heads. If the coins land on odds, all bets are frozen and the spinner tosses again until either heads or tails comes up. If five odds are tossed in a row all players lose.
a. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on heads.
b. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on tails.
c. What is the expected long-run profit (or loss) to the player?

LEARNING OBJECTIVE 3

Calculate average return and measure risk associated with various investment proposals

5.2 COVARIANCE AND ITS APPLICATION IN FINANCE

In Section 5.1 the expected value, variance and standard deviation of a discrete random variable are discussed. In this section the covariance between two discrete random variables is introduced and then applied to portfolio management, a topic of interest to financial analysts.

Covariance

covariance
Measure of the strength of the linear relationship between two numerical variables.

Covariance, σ_{XY}, is a measure of the strength of the relationship between two random variables, X and Y. A positive covariance indicates a positive relationship, while a negative covariance indicates a negative relationship. A covariance of zero indicates that the two variables are independent. Equation 5.4a defines the covariance between two discrete random variables.

COVARIANCE – DEFINITION FORMULA

$$\sigma_{XY} = \sum_{all\ X_i} \sum_{all\ Y_j} [X_i - E(X)][Y_j - E(Y)]P(X_i\ and\ Y_j) \qquad \textbf{(5.4a)}$$

where X_i is the ith outcome of the discrete random variable X, and Y_j is the jth outcome of the discrete random variable Y.

As for sample covariance, we can use algebra to obtain an alternative *calculation formula*.

COVARIANCE – CALCULATION FORMULA

$$\sigma_{XY} = \sum_{\text{all } X_i} \sum_{\text{all } Y_j} X_i Y_j P(X_i \text{ and } Y_j) - E(X)E(Y) \qquad \textbf{(5.4b)}$$

To illustrate covariance, suppose that we are deciding between two alternative investments for the coming year. The first investment is a mutual fund that consists of shares that are expected to do well when economic conditions are strong. The second investment is a mutual fund that is expected to perform best when economic conditions are weak. Your estimate of the returns for each investment (per $1,000 investment) under three economic conditions, each with a given probability of occurrence, is summarised in Table 5.3.

The expected value and standard deviation for each investment is calculated as follows.

Table 5.3 Estimated returns for each investment under three economic conditions

		Investment	
$P(X_iY_i)$	Economic condition	Strong-economy fund	Weak-economy fund
0.2	Recession	−$100	+$200
0.5	Stable economy	+100	+50
0.3	Expanding economy	+250	−100

Let X = strong-economy fund, and Y = weak-economy fund:

$$E(X) = \mu_X = (-100)(0.2) + (100)(0.5) + (250)(0.3) = \$105$$

$$E(Y) = \mu_Y = (+200)(0.2) + (50)(0.5) + (-100)(0.3) = \$35$$

$$\sigma^2_X = [(-100)^2 \times 0.2) + (100^2 \times 0.5) + (250^2 \times 0.3)] - 105^2 = 25{,}750 - 11{,}025 = 14{,}725$$

$$\sigma_X = \sqrt{\sigma^2_X} = \sqrt{14{,}725} = 121.346\ldots \approx \$121.35$$

$$\sigma^2_Y = [(200^2 \times 0.2) + (50^2 \times 0.5) + (-100)^2 \times 0.3)] - 35^2 = 12{,}250 - 1225 = 11{,}025$$

$$\sigma_Y = \sqrt{\sigma^2_Y} = \sqrt{11{,}025} = \$105.00$$

In the calculation of the covariance, the only non-zero probabilities are:

$$P(X = -\$100 \text{ and } Y = \$200) = 0.2$$
$$P(X = \$100 \text{ and } Y = \$50) = 0.5$$
$$P(X = \$250 \text{ and } Y = -\$100) = 0.3$$

We therefore have:

$$\sigma_{XY} = [(-100 \times 200 \times 0.2) + (100 \times 50 \times 0.5) + (250 \times (-100) \times 0.3)] - (105 \times 35)$$
$$= -9{,}000 - 3{,}675 = -12{,}675$$

Thus, the strong-economy fund has a higher expected value (i.e. larger expected return) than the weak-economy fund but has a higher standard deviation (i.e. more risk). The covariance of −12,675 between the two investments indicates a negative relationship in which the two investments are varying in the *opposite* direction. Therefore, when the return on one investment is high, the return on the other is typically low.

Expected Value, Variance and Standard Deviation of the Sum of Two Random Variables

Equation 5.4a defined the covariance between two discrete random variables, X and Y. Now, the **expected value of the sum of two random variables**, **variance of the sum of two random variables** and **standard deviation of the sum of two random variables** are defined.

expected value of the sum of two random variables
Measure of central tendency; mean of the sum of two random variables.

variance of the sum of two random variables
Measure of variation; directly related to the standard deviation.

standard deviation of the sum of two random variables
Measure of variation; directly related to the variance.

EXPECTED VALUE OF THE SUM OF TWO RANDOM VARIABLES

The expected value of the sum of two random variables is equal to the sum of the expected values.

$$E(X + Y) = E(X) + E(Y) \tag{5.5}$$

VARIANCE OF THE SUM OF TWO RANDOM VARIABLES

The variance of the sum of two random variables is equal to the sum of the variances plus twice the covariance.

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} \tag{5.6}$$

STANDARD DEVIATION OF THE SUM OF TWO RANDOM VARIABLES

The standard deviation is the square root of the variance.

$$\sigma_{X+Y} = \sqrt{\sigma^2_{X+Y}} \tag{5.7}$$

To illustrate the expected value, variance and standard deviation of the sum of two random variables, consider the two investments previously discussed. Using Equations 5.5, 5.6 and 5.7:

$$E(X + Y) = E(X) + E(Y) = 105 + 35 = \$140$$

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} = (14{,}725 + 11{,}025) + 2 \times (-12{,}675) = 400$$

$$\sigma_{X+Y} = \sqrt{400} = \$20$$

The expected return of the sum of the strong-economy fund and the weak-economy fund is \$140 with a standard deviation of \$20. The standard deviation of the sum of the two investments is much less than the standard deviation of either single investment because there is a large negative covariance between the investments.

Portfolio Expected Return and Portfolio Risk

portfolio
A combined investment in two or more assets.

portfolio expected return
Measure of central tendency; mean return on investment.

portfolio risk
Measure of the variation of investment returns.

The concepts of covariance, expected return and standard deviation of the sum of two random variables can be applied to the study of investment **portfolios** where investors combine assets into portfolios to reduce their risk. The objective is to maximise the return while minimising the risk. For such portfolios, rather than studying the sum of two random variables, each investment is weighted by the proportion of assets assigned to that investment. Equations 5.8 and 5.9 define **portfolio expected return** and **portfolio risk**.

PORTFOLIO EXPECTED RETURN

The portfolio expected return for a two-asset investment is equal to the weight assigned to asset X multiplied by the expected return of asset X plus the weight assigned to asset Y multiplied by the expected return of asset Y:

$$E(P) = wE(X) + (1 - w)E(Y) \tag{5.8}$$

where $E(P)$ = portfolio expected return
w = portion of the portfolio assigned to asset X, $0 \leq w \leq 1$
$1 - w$ = portion of the portfolio assigned to asset Y

PORTFOLIO RISK

$$\sigma_p = \sqrt{w^2\sigma^2_X + (1-w)^2\sigma^2_Y + 2w(1-w)\sigma_{XY}} \tag{5.9}$$

In the previous example, the expected return and risk of two different investments were evaluated, a strong-economy fund and a weak-economy fund. The covariance of the two investments was also calculated. Now, suppose that we wish to form a portfolio of these two investments that consists of an equal investment in each of these two funds. To calculate the portfolio expected return and the portfolio risk, use Equations 5.8 and 5.9, with $w = 0.5$, to obtain:

$$E(P) = wE(X) + (1-w)E(Y) = (0.5 \times 105) + (0.5 \times 35) = \$70$$

$$\begin{aligned}\sigma_p &= \sqrt{(0.5)^2(14{,}725) + (1-0.5)^2(11{,}025) + 2(0.5)(1-0.5)(-12{,}675)}\\ &= \sqrt{100}\\ &= \$10\end{aligned}$$

Thus, the portfolio has an expected return of \$70 for each \$1,000 invested (a return of 7%) and has a portfolio risk of \$10. The portfolio risk here is small because there is a large negative covariance between the two investments. The fact that each investment performs best under different circumstances has reduced the overall risk of the portfolio.

Problems for Section 5.2

LEARNING THE BASICS

5.6 Given the following probability distributions for variables X and Y:

$P(X_iY_i)$	X	Y
0.4	100	200
0.6	200	100

Calculate:

a. $E(X)$ and $E(Y)$
b. σ_X and σ_Y
c. σ_{XY}
d. $E(X + Y)$

5.7 Given the following probability distributions for variables X and Y:

$P(X_iY_i)$	X	Y
0.2	−100	50
0.4	50	30
0.3	200	20
0.1	300	20

Calculate:

a. $E(X)$ and $E(Y)$
b. σ_X and σ_Y
c. σ_{XY}
d. $E(X + Y)$

5.8 Two investments, X and Y, have the following characteristics:

$E(X) = \$50$, $E(Y) = \$100$, $\sigma_X^2 = 9{,}000$, $\sigma_Y^2 = 15{,}000$ and $\sigma_{XY} = 7{,}500$

If the weight assigned to investment X of portfolio assets is 0.4, calculate:

a. the portfolio expected return
b. the portfolio risk

APPLYING THE CONCEPTS

5.9 The process of being served at a bank consists of two independent parts – the time waiting in line and the time it takes to be served by the teller. Suppose, at a branch of Check\$mart, that the time waiting in line has an expected value of 4 minutes with a standard deviation of 1.2 minutes and the time it takes to be served by the teller has an expected value of 5.5 minutes with a standard deviation of 1.5 minutes. Calculate:

a. the expected value of the total time it takes to be served
b. the standard deviation of the total time it takes to be served

5.10 For the investment example given in Table 5.3:

a. Calculate the portfolio expected return and the portfolio risk if:
  i. 30% is invested in the strong-economy fund and 70% in the weak-economy fund
  ii. 70% is invested in the strong-economy fund and 30% in the weak-economy fund

b. Which of the three investment strategies (30%, 50% or 70% in the strong-economy fund) would you recommend? Why?

5.11 You are trying to develop a strategy for investing in two different shares. The anticipated annual return for a $1,000 investment in each share has the following probability distribution:

	Returns	
Probability	**Share X**	**Share Y**
0.1	−$100	$50
0.3	0	150
0.3	80	−20
0.3	150	−100

a. Calculate:
- **i.** the expected return for share *X* and for share *Y*
- **ii.** the standard deviation for share *X* and for share *Y*
- **iii.** the covariance of share *X* and share *Y*

b. Would you invest in share *X* or share *Y*? Explain.

5.12 Suppose that in problem 5.11 you wanted to create a portfolio that consists of share *X* and share *Y*.

a. Calculate the portfolio expected return and portfolio risk for each of the following percentages invested in share *X*:
- **i.** 30%
- **ii.** 50%
- **iii.** 70%

b. On the basis of the results of your calculations in part (a), which portfolio would you recommend? Explain.

5.13 You are trying to set up a portfolio that consists of a corporate bond fund and a common share fund. The following information about the annual return (per $1,000) of each of these investments under different economic conditions is available, together with the probability that each of these economic conditions will occur.

Probability	State of the economy	Corporate bond fund	Common share fund
0.10	Recession	−$30	−$150
0.15	Stagnation	50	−20
0.35	Slow growth	90	120
0.30	Moderate growth	100	160
0.10	High growth	110	250

a. Calculate:
- **i.** the expected return for the corporate bond fund and for the common share fund
- **ii.** the standard deviation for the corporate bond fund and for the common share fund
- **iii.** the covariance of the corporate bond fund and the common share fund

b. Would you invest in the corporate bond fund or the common share fund? Explain.

5.14 Suppose that in problem 5.13 you wanted to create a portfolio that consists of a corporate bond fund and a common share fund.

a. Calculate the portfolio expected return and portfolio risk for each of the following percentages invested in a corporate bond fund:
- **i.** 30%
- **ii.** 50%
- **iii.** 70%

b. On the basis of the results of your calculations in (a), which portfolio would you recommend? Explain.

5.3 BINOMIAL DISTRIBUTION

LEARNING OBJECTIVE 4

Identify situations that can be modelled by a binomial distribution and calculate binomial probabilities

mathematical model
The mathematical representation of a random variable.

binomial distribution
Discrete probability distribution, where the random variable is the number of successes in a sample of *n* observations from either an infinite population or sampling with replacement.

The next three sections use mathematical models to solve business and other problems.

> A **mathematical model** is a mathematical expression representing a variable of interest.

When a mathematical model of a discrete probability distribution is available, you can easily calculate the exact probability of occurrence of any particular outcome of the random variable.

The **binomial distribution** is one of the most important and widely used discrete probability distributions. The binomial distribution arises when the discrete random variable is the number of successes in a sample of n observations. The binomial distribution has four essential properties:

1. The sample consists of a fixed number of observations, n.
2. Each observation is classified into one of two mutually exclusive and collectively exhaustive categories, usually called *success* and *failure*.
3. The probability of an observation being classified as success, p, is constant from observation to observation. Thus, the probability of an observation being classified as failure, $1 - p$, is also constant for all observations.
4. The outcome (i.e. success or failure) of any observation is independent of the outcome of any other observation. To ensure independence, the observations can be randomly selected either from an *infinite population without replacement* or from a *finite population with replacement*.

The proportion of enquiries that are converted to confirmed bookings is of interest in the Gaia Cruises scenario, so Yang could define an enquiry converted to a booking as a *success* and an enquiry that is not converted to a booking as a *failure*. Yang would then be interested in the number of successes; that is, the number of enquiries converted to bookings in a random sample of n online enquiries. *Note:* In a binomial distribution, 'success' is usually defined as the outcome we are interested in – in this case converted enquiries.

This is a binomial situation because:

- a fixed number of online enquiries, n, is chosen
- each online enquiry is either converted to a booking – a success – or not converted – a failure
- 10% of enquiries are converted to bookings, so the probability of a randomly chosen enquiry being converted to a booking is $p = 0.1$ and that of a randomly chosen enquiry not being converted to a booking is $1 - p = 0.9$
- online enquiries are independent and randomly selected; so the outcome, converted or not converted, of any enquiry is independent of the outcome of any other enquiry.

If Yang takes a random sample of four online enquiries, the binomial random variable defined as:

$$X = \text{number of converted enquiries}$$

has a range from 0 to four as none, one, two, three or all four enquiries may be converted to bookings. In general, a binomial random variable has a range from 0 to n.

Suppose that Yang observes the following result in a sample of four enquiries:

First order	**Second order**	**Third order**	**Fourth order**
Converted	Converted	Not converted	Converted

What is the probability of having three successes (converted enquiries) in a sample of four enquiries in this particular sequence? Because the historical probability of enquiries converted to bookings is 0.10, the probability that each enquiry occurs in the sequence is:

First enquiry	**Second enquiry**	**Third enquiry**	**Fourth enquiry**
$p = 0.1$	$p = 0.1$	$1 - p = 0.9$	$p = 0.1$

Each outcome is independent of the others because the enquiries are independent and are selected randomly. Therefore, the probability of having this particular sequence is:

$$pp(1-p)p = p^3(1-p) = (0.1)^3(0.9)^1 = 0.0009$$

This result indicates only the probability of three enquiries converted to bookings (successes) out of a sample of four online enquiries in a *specific sequence*. To find the number of ways of selecting X objects out of n objects *irrespective of sequence*, use the counting rule for combinations introduced in Chapter 4 as Equation 4.14 and given below, introducing a different notation, as Equation 5.10.

COMBINATIONS

The number of combinations of selecting X objects out of n objects is given by

$$\binom{n}{X} = {}_nC_X = \frac{n!}{X!(n-X)!} \tag{5.10}$$

where n factorial is defined by $n! = n \times (n-1) \times \ldots \times 2 \times 1$ and by definition, $0! = 1$.

Using Equation 5.10 we see that there are:

$${}_4C_3 = \frac{4!}{3!(4-3)!} = 4$$

sequences of three converted enquiries and one enquiry not converted. The four possible sequences are:

Sequence 1	Converted	Converted	Converted	Not converted
Sequence 2	Converted	Converted	Not converted	Converted
Sequence 3	Converted	Not converted	Converted	Converted
Sequence 4	Not converted	Converted	Converted	Converted

and the probability of each is:

$$p^3(1-p) = (0.1)^3(0.9)^1 = 0.0009$$

Therefore, the probability of three converted enquiries out of four is equal to:

$$\text{number of sequences} \times \text{probability of sequence} = 4 \times 0.0009 = 0.0036$$

We can make similar, intuitive derivations for the other possible outcomes of the random variable – zero, one, two and four converted enquiries. However, as n, the sample size, gets larger, the calculations involved in using this approach become time-consuming. Instead, a mathematical model provides a formula to calculate any binomial probability. Equation 5.11 is the mathematical model that represents the binomial probability distribution and is used to calculate the probability of X successes for any given values of n and p.

BINOMIAL PROBABILITY DISTRIBUTION

$$P(X) = \frac{n!}{X!(n-X)!}p^X(1-p)^{n-X} \tag{5.11}$$

where $P(X)$ = probability of X successes given n and p
n = number of observations
p = probability of success
$1-p$ = probability of failure
X = number of successes ($X = 0, 1, 2, \dots, n$)

Equation 5.11 restates what we had intuitively derived. The binomial random variable X can have any integer value X from 0 to n. In Equation 5.11 the product:

$$p^X(1-p)^{n-X}$$

indicates the probability of exactly X successes out of n observations in a *particular sequence*. The term:

$$\frac{n!}{X!(n-X)!}$$

indicates *how many combinations* of the X successes out of n observations are possible.

Hence, given the number of observations n and the probability of success p, the probability of X successes is:

$$\begin{aligned} P(X) &= \text{number of sequences} \times \text{probability of sequence} \\ &= \frac{n!}{X!(n-X)!}p^X(1-p)^{n-X} \end{aligned}$$

Example 5.1 illustrates the use of Equation 5.11.

EXAMPLE 5.1 DETERMINING $P(X = 3)$, GIVEN $n = 4$ AND $p = 0.1$

If 10% of online enquiries are converted to bookings, what is the probability that there are three converted enquiries in a sample of four?

SOLUTION

Using Equation 5.11, the probability of three converted enquiries from a sample of four is:

$$P(X = 3) = \frac{4!}{3!(4-3)!}(0.1)^3(1-0.1)^{4-3} = 4 \times 0.001 \times 0.9 = 0.0036$$

Examples 5.2 and 5.3 give the calculations for other values of X.

EXAMPLE 5.2

DETERMINING $P(X \geqslant 3)$, GIVEN $n = 4$ AND $p = 0.1$

If 10% of online enquiries are converted to bookings, what is the probability that there are at least three converted enquiries in a sample of four?

SOLUTION

In Example 5.1 we found that the probability of *exactly* three converted enquiries from a sample of four is 0.0036. To calculate the probability of *at least* three converted enquiries, we need to add the probability of three converted enquiries to the probability of four converted enquiries. The probability of four converted enquiries is:

$$P(X = 4) = \frac{4!}{4!(4-4)!}(0.1)^4(1-0.1)^{4-4} = 1 \times 0.0001 \times 1 = 0.0001$$

Thus, the probability of at least three converted enquiries is:

$$P(X \geqslant 3) = P(X = 3) + P(X = 4) = 0.0036 + 0.0001 = 0.0037$$

There is a 0.37% chance that there will be at least three converted enquiries in a sample of four.

EXAMPLE 5.3

DETERMINING $P(X < 3)$, GIVEN $n = 4$ AND $p = 0.1$

If 10% of online enquiries are converted to bookings, what is the probability that there are fewer than three converted enquiries in a sample of four?

SOLUTION

The probability that there are fewer than three converted enquiries is:

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

Use Equation 5.11 to calculate each of these probabilities:

$$P(X = 0) = \frac{4!}{0!(4-0)!}(0.1)^0(1-0.1)^{4-0} = 0.6561$$

$$P(X = 1) = \frac{4!}{1!(4-1)!}(0.1)^1(1-0.1)^{4-1} = 0.2916$$

$$P(X = 2) = \frac{4!}{2!(4-2)!}(0.1)^2(1-0.1)^{4-2} = 0.0486$$

Therefore, $P(X < 3) = 0.6561 + 0.2916 + 0.0486 = 0.9963$

Alternatively, $P(X < 3)$ can also be calculated from its complement, $P(X \geqslant 3)$, since:

$$P(X < 3) = 1 - P(X \geqslant 3) = 1 - 0.0037 = 0.9963$$

Calculations such as those in Example 5.3 can become tedious, especially as n gets large. To avoid computational drudgery, many binomial probabilities can be found directly from Table E.6 (Appendix E), a portion of which is reproduced in Table 5.4. Table E.6 provides binomial probabilities for $X = 0, 1, 2, \ldots, n$ for selected combinations of n and p. For example, to find the probability of exactly two successes in a sample of four when the probability of success is 0.1, first find $n = 4$ and then read off the required probability at the intersection of the row $X = 2$ and the column $p = 0.10$. Thus:

$$P(X = 2) = 0.0486$$

Table 5.4 Finding a binomial probability for $n = 4$, $X = 2$ and $p = 0.1$ (extracted from Table E.6)

				p	
n	***X***	**0.01**	**0.02**	**....**	**0.10**
4	0	0.9606	0.9224	....	0.6561
	1	0.0388	0.0753	....	0.2916
	2	0.0006	0.0023	....	0.0486
	3	0.0000	0.0000	....	0.0036
	4	0.0000	0.0000	....	0.0001

The binomial probabilities given in Table E.6 can also be calculated using Microsoft Excel. Figure 5.2 presents a Microsoft Excel worksheet for calculating binomial probabilities, using the Excel 2010 and later inbuilt binomial function BINOM.DIST(number_s,trials, probability_s,cumulative). For earlier versions of Excel the corresponding binomial function is BINOMDIST(number_s,trials,probability_s,cumulative).

Figure 5.2 Microsoft Excel worksheet for calculating binomial probabilities

	A	B	C
3	**Data**		
4	Sample size	4	
5	Probability of success	0.1	
6			
7	**Statistics**		
8	Mean	0.4	=B4*B5
9	Variance	0.36	=B8*(1-B5)
10	Standard deviation	0.6	=SQRT(B9)
11			
12	Binomial probabilities table		
13	*X*	*P*(*X*)	
14	0	0.6561	=BINOM.DIST(A14,B4,B5,FALSE)
15	1	0.2916	=BINOM.DIST(A15,B4,B5,FALSE)
16	2	0.0486	=BINOM.DIST(A16,B4,B5,FALSE)
17	3	0.0036	=BINOM.DIST(A17,B4,B5,FALSE)
18	4	0.0001	=BINOM.DIST(A18,B4,B5,FALSE)

The shape of a binomial probability distribution depends on the values of n and p. When $p = 0.5$, the binomial distribution is symmetrical, regardless of how large or small the value of n. When $p \neq 0.5$, the distribution is skewed, to the right if $p < 0.5$ and to the left if $p > 0.5$. The closer p is to 0.5 and the larger the number of observations n, the less skewed the distribution. For example, the distribution of the number of converted enquiries is highly skewed to the right because $p = 0.1$ and $n = 4$ (see Figure 5.3).

Substituting the binomial probability equation (5.11) in the expected value equation (5.1) and using algebra to simplify, it can be shown that the mean of the binomial distribution is equal to the product of n and p, as shown in Equation 5.12. Therefore, use Equation 5.12 to calculate the mean of a binomial distribution, instead of Equation 5.1.

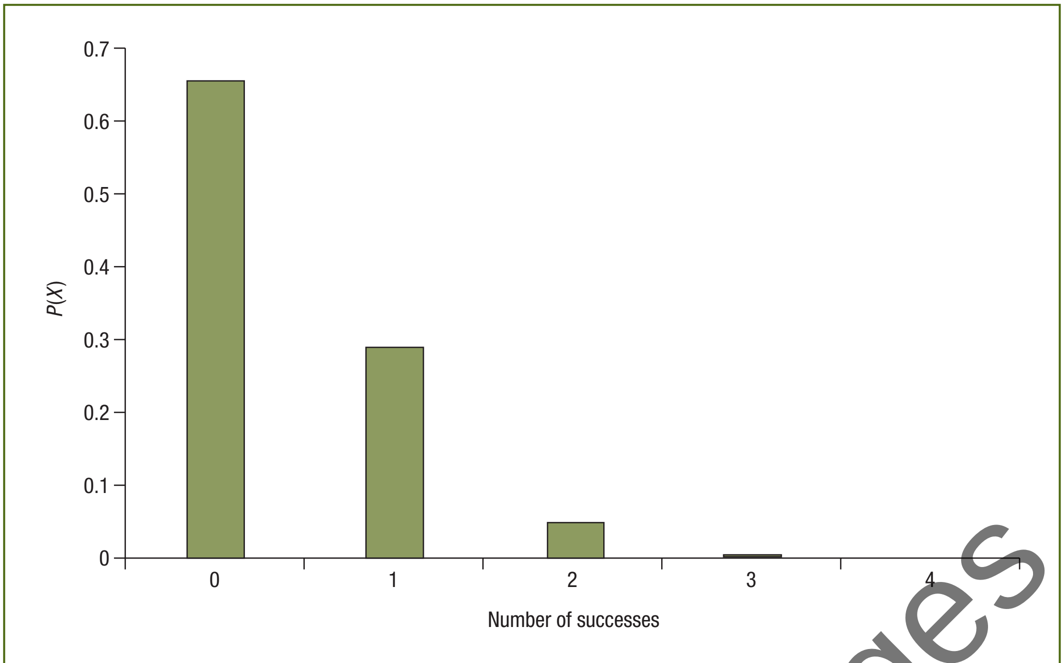

Figure 5.3
Microsoft Excel graph of the binomial probability distribution with $n = 4$ and $p = 0.1$

THE MEAN OF THE BINOMIAL DISTRIBUTION

The mean μ of the binomial distribution is equal to the sample size n multiplied by the probability of success p.

$$\mu = E(X) = np \tag{5.12}$$

Therefore, on average, Yang can theoretically expect $E(X) = 4 \times 0.1 = 0.4$ converted enquiries in a sample of four.

Similarly, by substituting the binomial probability equation (5.11) in the variance equation (5.2a or 5.2b) and using algebra to simplify, it can be shown that the standard deviation of the binomial distribution is given by Equation 5.13.

THE STANDARD DEVIATION OF THE BINOMIAL DISTRIBUTION

$$\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)} \tag{5.13}$$

Therefore, using Equation 5.13, the standard deviation of the number of converted enquiries is:

$$\sigma = \sqrt{4(0.1)(0.9)} = 0.60$$

EXAMPLE 5.4

CALCULATING BINOMIAL PROBABILITIES

Accuracy (measured as the percentage of orders consisting of a main item, side item and drink that are filled correctly) in taking orders at the drive-through window is an important feature for fast-food chains. Suppose in a recent month that records show that the percentage of correct orders of this type filled at a Hungry Jack's franchise was 88%. Suppose you and two friends go to the drive-through window at this Hungry Jack's franchise and each of you places an order of the type just mentioned:

- What is the probability that:
  - all three orders will be filled correctly?
  - none of the three will be filled correctly?
  - at least two of the three will be filled correctly?
- What is the average and standard deviation of the number of orders filled correctly?

SOLUTION

There are three orders and the probability of any order being accurate is 0.88. Therefore:

$$X = \text{number of orders filled correctly} = 0, 1, 2, 3$$

is a binomial random variable with $n = 3$, $p = 0.88$.

Using Equations 5.11, 5.12 and 5.13:

$$P(X = 3) = \frac{3!}{3!(3-3)!}(0.88)^3(1-0.88)^{3-3} = 1 \times 0.68147\ldots \times 1 = 0.68147\ldots$$

$$P(X = 0) = \frac{3!}{0!(3-0)!}(0.88)^0(1-0.88)^{3-0} = 1 \times 1 \times 0.00172\ldots = 0.00172$$

$$P(X = 2) = \frac{3!}{2!(3-2)!}(0.88)^2(1-0.88)^{3-2} = 3 \times 0.7744 \times 0.12\ldots = 0.27878\ldots$$

$$P(X \geqslant 2) = P(X = 2) + P(X = 3) = 0.27878\ldots + 0.68147\ldots = 0.96025\ldots$$

$$\mu = E(X) = 3 \times 0.88 = 2.64$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{3 \times 0.88 \times 0.12} = 0.5628\ldots$$

The probability that all three orders are filled correctly is 0.6815. The probability that none of the orders is filled correctly is 0.0017. The probability that at least two orders are filled correctly is 0.9603. The mean number of accurate orders filled in a sample of three orders is 2.64 and the standard deviation is 0.563.

This section introduced the binomial distribution and applied it to business and other problems. The binomial distribution plays an important role when it is used in statistical inference problems involving the estimation or testing of hypotheses about proportions (discussed in Chapters 8 and 9).

Problems for Section 5.3

Problems 5.15 to 5.24 can be solved manually or by using Microsoft Excel; some, but not all, can also be solved using Table E.6.

LEARNING THE BASICS

5.15 If X is a binomial random variable, determine the following:

a. For $n = 4$ and $p = 0.12$, what is $P(X = 0)$?
b. For $n = 10$ and $p = 0.40$, what is $P(X = 9)$?
c. For $n = 10$ and $p = 0.50$, what is $P(X = 8)$?
d. For $n = 6$ and $p = 0.83$, what is $P(X = 5)$?

5.16 If X is a binomial random variable with $n = 5$ and $p = 0.40$, what is the probability that:

a. $X = 4$?
b. $X \leqslant 3$?
c. $X < 2$?
d. $X > 1$?

5.17 Determine the mean and standard deviation of the random variable X in each of the following binomial distributions:

a. $n = 4$ and $p = 0.10$
b. $n = 4$ and $p = 0.40$
c. $n = 5$ and $p = 0.80$
d. $n = 3$ and $p = 0.50$

APPLYING THE CONCEPTS

5.18 The increase or decrease in the price of a share between the beginning and the end of a trading day is assumed to be an equally likely random event. What is the probability that a share will show an increase in its closing price on five consecutive days?

5.19 Research has shown that only 60% of consumers read every word, including the fine print, of a service contract. Assume that the number of consumers who read every word of a contract can be modelled using the binomial distribution. A group of five consumers has just signed a 12-month contract with an ISP (Internet service provider).

a. What is the probability that:

i. all five will have read every word of their contract?

ii. at least three will have read every word of their contract?

iii. less than two will have read every word of their contract?

b. What would your answers be in (a) if the probability is 0.80 that a consumer reads every word of a service contract?

5.20 A student taking a multiple-choice exam with five questions in which each question has four options selects the answers randomly. What is the probability that the student will get:

a. five questions correct?

b. at least four questions correct?

c. no questions correct?

d. no more than two questions correct?

5.21 In Example 5.4 you and two friends went to a Hungry Jack's franchise. Instead, suppose that you go to a McDonald's franchise, which last month filled 90% of orders correctly.

a. What is the probability that:

i. all three orders will be filled correctly?

ii. none of the three will be filled correctly?

iii. at least two of the three will be filled correctly?

b. What is the mean and standard deviation of the number of orders filled correctly?

5.22 In a certain weekday television show, the winning contestant has to choose randomly from 20 boxes, one of which contains a major prize of $100,000.

a. What is the probability that, during a week:

i. no contestant wins the major prize?

ii. exactly one contestant wins the major prize?

iii. no more than two contestants win the major prize?

iv. at least three contestants win the major prize?

b. Calculate the expected number and standard deviation of winners in a week.

c. How much should the producers budget for major prizes per week?

5.23 When a customer places an order with Rudy's On-Line Office Supplies, a computerised accounting information system (AIS) automatically checks to see whether the customer has exceeded their credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable.

a. What are the mean and standard deviation of the number of customers exceeding their credit limits?

b. What is the probability that no customer will exceed their limit?

c. What is the probability that one customer will exceed their limit?

d. What is the probability that two or more customers will exceed their limits?

5.24 A new drug is found to be effective on 90% of the patients tested.

a. Is the 90% effective rate best classified as *a priori* classical probability, empirical classical probability or subjective probability?

b. If the drug is administered to 20 randomly chosen patients at a large hospital, find the probability that it is effective for:

i. fewer than five of the patients

ii. 10 or more of the patients

iii. all 20 of the patients

5.4 POISSON DISTRIBUTION

LEARNING OBJECTIVE 5

Identify situations that can be modelled by a Poisson distribution and calculate Poisson probabilities

Many studies are based on the number of times a random event occurs in an interval of time or space. Examples are the number of surface defects on a new refrigerator, the number of network failures in a day or the number of fleas on the body of a dog.

The **Poisson distribution** can be used to calculate probabilities when counting the number of times a particular event occurs in an interval of time or space if:

1. the probability an event occurs in any interval is the same for all intervals of the same size
2. the number of occurrences of the event in one non-overlapping interval is independent of the number in any other interval
3. the probability that two or more occurrences of the event in an interval approaches zero as the interval becomes smaller.

Poisson distribution
Discrete probability distribution, where the random variable is the number of events in a given interval.

If these properties hold, then the average or expected number of occurrences over any interval is proportional to the size of the interval.

Consider the number of online enquiries received. Suppose that Yang is interested in the number of online enquiries received in 8 minutes. Does this situation match the properties of the Poisson distribution given above? First, define the random variable as:

$$X = \text{number of online enquiries received in 8 minutes}$$

Suppose enquiries are received randomly, then it is reasonable to assume that the probability that an enquiry is received during an 8-minute interval is the same as the probability for all the other 8-minute intervals. Yang can also assume that the receipt of an enquiry during an 8-minute interval has no effect on (i.e. is statistically independent of) the receipt of any other enquiry during any 8-minute interval. Finally, the probability that two or more enquiries will be received in a given time period approaches zero as the time interval becomes smaller. For example, the probability is virtually zero that two enquiries will be received in a time interval of 0.001 of a second. Thus, Yang can use the Poisson distribution to determine probabilities involving the number of enquiries received in an 8-minute interval.

The Poisson distribution has one parameter, λ (the Greek lower-case letter *lambda*), which is the mean or expected number of events per interval. The variance of a Poisson distribution is also equal to λ, hence the standard deviation is equal to $\sqrt{\lambda}$. The number of events, X, of the Poisson random variable ranges from 0 to infinity.

Equation 5.14, the mathematical formula for the Poisson distribution, gives the probability of X events in an interval, given that λ events are expected.

POISSON PROBABILITY DISTRIBUTION

$$P(X) = \frac{e^{-\lambda}\lambda^X}{X!} \tag{5.14}$$

where $P(X)$ = the probability of X events in a given interval
λ = expected number of events in the given interval
e = 2.71828 … is the base of natural logarithms

To illustrate the use of the Poisson distribution, calculate the probability that in a given 8 minutes exactly five online enquiries will be received, and the probability that less than five online enquiries will be received. As, on average, Gaia Cruises receives 75 online enquiries an hour, the average or expected number of enquiries per 8 minutes is $\lambda = \frac{75}{60} \times 8 = 10$

Using Equation 5.14 with $\lambda = 10$, the probability that in a given 8 minutes exactly five online enquiries will be received is:

$$P(X = 5) = \frac{e^{-10}10^5}{5!} = \frac{4.53999\ldots}{120} = 0.03783\ldots$$

and the probability that in any given 8 minutes less than five online enquiries will be received is:

$$\begin{aligned} P(X < 5) &= \frac{e^{-10}(10)^0}{0!} + \frac{e^{-10}(10)^1}{1!} + \frac{e^{-10}(10)^2}{2!} + \frac{e^{-10}(10)^3}{3!} + \frac{e^{-10}(10)^4}{4!} \\ &= 0.00004\ldots + 0.00045\ldots + 0.00226\ldots + 0.00756\ldots + 0.01891\ldots \\ &= 0.029252\ldots \end{aligned}$$

Thus, there is a 3% likelihood that less than five online enquiries are received in 8 minutes, leading to enquiry staff having significant idle time.

To avoid the computational drudgery involved in these calculations, many Poisson probabilities can be found directly from Table E.7 (Appendix E), a portion of which is reproduced in

Table 5.5. Table E.7 provides the probabilities that the Poisson random variable takes on values of $X = 0, 1, 2, \ldots$ for selected values of the parameter λ. The probability that exactly five enquiries will be received in a given 8-minute interval when the mean number of enquiries received in 8 minutes is 10 is given by the intersection of the row $X = 5$ and column $\lambda = 10$. Therefore, from Table 5.5, $P(X = 5) = 0.0378$.

Table 5.5
Calculating a Poisson probability for $\lambda = 10$ (extracted from Table E.7 in Appendix E of this book)

		λ		
X	**9.1**	**9.2**		**10**
0	0.0001	0.0001	....	0.0000
1	0.0010	0.0009	....	0.0005
2	0.0046	0.0043	....	0.0023
3	0.0140	0.0131	....	0.0076
4	0.0319	0.0302	....	0.0189
5	0.0581	0.0555	....	0.0378
6	0.0881	0.0851	....	0.0631
7	0.1145	0.1118	....	0.0901

You can also calculate the Poisson probabilities given in Table E.7 using Microsoft Excel. Figure 5.4 presents a Microsoft Excel worksheet for the Poisson distribution, with $\lambda = 10$, using the Excel 2010 and later inbuilt Poisson function POISSON.DIST(x,mean,cumulative). For earlier versions of Excel the corresponding Poisson function is POISSON(x,mean,cumulative).

	A	B	C	D	E
3			**Data**		
4	Average/expected number of successes:				10
5					
6	Poisson probabilities table				
7	*X*	*P(X)*			
8	0	0.000045	=POISSON.DIST($A8,$E$4,FALSE)		
9	1	0.000454	=POISSON.DIST($A9,$E$4,FALSE)		
10	2	0.002270	=POISSON.DIST($A10,$E$4,FALSE)		
11	3	0.007567	=POISSON.DIST($A11,$E$4,FALSE)		
12	4	0.018917	=POISSON.DIST($A12,$E$4,FALSE)		
13	5	0.037833	=POISSON.DIST($A13,$E$4,FALSE)		
14	6	0.063055	=POISSON.DIST($A14,$E$4,FALSE)		
15	7	0.090079	=POISSON.DIST($A15,$E$4,FALSE)		
16	8	0.112599	=POISSON.DIST($A16,$E$4,FALSE)		
17	9	0.125110	=POISSON.DIST($A17,$E$4,FALSE)		
18	10	0.125110	=POISSON.DIST($A18,$E$4,FALSE)		
19	11	0.113736	=POISSON.DIST($A19,$E$4,FALSE)		
20	12	0.094780	=POISSON.DIST($A20,$E$4,FALSE)		
21	13	0.072908	=POISSON.DIST($A21,$E$4,FALSE)		
22	14	0.052077	=POISSON.DIST($A22,$E$4,FALSE)		
23	15	0.034718	=POISSON.DIST($A23,$E$4,FALSE)		
24	16	0.021699	=POISSON.DIST($A24,$E$4,FALSE)		
25	17	0.012764	=POISSON.DIST($A25,$E$4,FALSE)		
26	18	0.007091	=POISSON.DIST($A26,$E$4,FALSE)		
27	19	0.003732	=POISSON.DIST($A27,$E$4,FALSE)		
28	20	0.001866	=POISSON.DIST($A28,$E$4,FALSE)		

Figure 5.4
Microsoft Excel worksheet for calculating Poisson probabilities

EXAMPLE 5.5

CALCULATING POISSON PROBABILITIES

The number of faults per month that arise in the gearboxes of a bus fleet is known to follow a Poisson distribution with a mean of 2.5 faults per month. What is the probability that in a given month no faults are found? At least one fault is found?

SOLUTION

Using Equation 5.14 with $\lambda = 2.5$ (or using Table E.7 or Microsoft Excel), the probabilities that in a given month no faults are found and at least one fault is found are:

$$P(X = 0) = \frac{e^{-2.5}(2.5)^0}{0\,!} = \frac{0.08208\ldots \times 1}{1} = 0.08208\ldots$$

$$P(X \geqslant 1) = 1 - P(X = 0) = 1 - 0.08208\ldots = 0.91791\ldots$$

The probability that there will be no faults in a given month is 0.0821. The probability that there will be at least one fault is 0.9179.

EXAMPLE 5.6

CALCULATING POISSON PROBABILITIES

For the Gaia Cruises scenario, what is the probability that 24 or more online enquiries are received in 12 minutes?

SOLUTION

Let X = number of enquiries received in 12 minutes, then X is Poisson with $\lambda = \frac{75}{60} \times 12 = 15$

Using Microsoft Excel we can obtain Table 5.6, which gives Poisson probabilities for $\lambda = 15$.

Table 5.6
Poisson probabilities for $\lambda = 15$

Enquiries received in 12 minutes
Expected number of enquiries: 15

X	*P*(*X*)	*X*	*P*(*X*)	*X*	*P*(*X*)
0	0.0000	8	0.0194	16	0.0960
1	0.0000	9	0.0324	17	0.0847
2	0.0000	10	0.0486	18	0.0706
3	0.0002	11	0.0663	19	0.0557
4	0.0006	12	0.0829	20	0.0418
5	0.0019	13	0.0956	21	0.0299
6	0.0048	14	0.1024	22	0.0204
7	0.0104	15	0.1024	23	0.0133
				Total	0.9805

From Table 5.6:

$$P(X \geqslant 24) = 1 - P(X < 24) = 1 - P(X \leqslant 23) = 1 - 0.9805 = 0.0195$$

Therefore, in approximately 2% of 12-minute intervals 24 or more enquiries are received, hence increasing the likelihood that enquiries start to queue and may not be answered within the stated 2 hours.

Problems for Section 5.4

LEARNING THE BASICS

5.25 Assume a Poisson distribution.
- **a.** If $\lambda = 2.5$, find $P(X = 2)$.
- **b.** If $\lambda = 8$, find $P(X = 8)$.
- **c.** If $\lambda = 0.5$, find $P(X = 1)$.
- **d.** If $\lambda = 3.7$, find $P(X = 0)$.

5.26 Assume a Poisson distribution.
- **a.** If $\lambda = 2$, find $P(X \geqslant 2)$.
- **b.** If $\lambda = 8$, find $P(X \geqslant 3)$.
- **c.** If $\lambda = 0.5$, find $P(X \leqslant 1)$.
- **d.** If $\lambda = 4$, find $P(X \geqslant 1)$.
- **e.** If $\lambda = 5$, find $P(X \leqslant 3)$.

5.27 Assume a Poisson distribution with $\lambda = 5$. Find the probability that:
- **a.** $X = 1$
- **b.** $X < 1$
- **c.** $X > 1$
- **d.** $X \leqslant 1$

APPLYING THE CONCEPTS

Problems 5.28 to 5.32 can be solved manually or by using Microsoft Excel. Some, but not all, can also be solved using Table E.7.

5.28 The quality control manager of Marilyn's Bakery is inspecting a batch of chocolate-chip biscuits that has just been baked. If the production process is in control, the mean number of chip parts per biscuit is 6.0. What is the probability that, in any particular biscuit being inspected:
- **a.** fewer than five chip parts will be found?
- **b.** exactly five chip parts will be found?
- **c.** five or more chip parts will be found?
- **d.** either four or five chip parts will be found?

5.29 Refer to problem 5.28. How many biscuits in a batch of 100 should the manager expect to discard if company policy requires that all chocolate-chip biscuits sold must have at least four chocolate-chip parts?

5.30 The number of floods in a certain region is approximately Poisson distributed with an average of three floods every 10 years.
- **a.** Find the probability that a family living in the area for one year will experience:
  - **i.** exactly one flood
  - **ii.** at least one flood
- **b.** Find the probability that a student who moves to the area for three years will experience
  - **i.** exactly one flood
  - **ii.** at least one flood

5.31 Based on past experience, it is assumed that the number of flaws per metre in rolls of grade 2 paper follow a Poisson distribution with a mean of one flaw per 5 metres of paper. What is the probability that in a:
- **a.** 1-metre roll there will be at least two flaws?
- **b.** 10-metre roll there will be at least one flaw?
- **c.** 50-metre roll there will be between five and 15 (inclusive) flaws?

5.32 A toll-free phone number is available from 9 am to 9 pm for customers to register a complaint about a product purchased from a large company. Past history indicates that an average of 0.4 calls are received per minute.
- **a.** What properties must be true about the situation described above in order to use the Poisson distribution to calculate probabilities concerning the number of phone calls received in a 1-minute period?
- **b.** Assuming that this situation matches the properties you discuss in (a), what is the probability that, during a 1-minute period:
  - **i.** zero phone calls will be received?
  - **ii.** three or more phone calls will be received?
- **c.** What is the maximum number of phone calls that will be received in a 1-minute period 99.99% of the time?

LEARNING OBJECTIVE 6

Identify situations that can be modelled by a hypergeometric distribution and calculate hypergeometric probabilities

5.5 HYPERGEOMETRIC DISTRIBUTION

The binomial distribution and the **hypergeometric distribution** are both concerned with the number of successes in a sample of n observations. However, they differ in the way in which the sample is selected. For the binomial distribution, as the probability of success p must be constant for all observations and the outcome of any particular observation must be independent of any other, the random sample is either selected *with* replacement from a *finite* population or *without* replacement from an *infinite* population. For the hypergeometric distribution, the random sample is selected *without* replacement from a *finite* population. Thus, the outcome of one observation is dependent on the outcomes of previous observations.

Consider a population of size N. Let A represent the total number of successes in the population. The hypergeometric distribution is then used to find the probability of X successes in a

hypergeometric distribution
Discrete probability distribution where the random variable is the number of successes in a sample of n observations from a finite population without replacement.

sample of size n selected without replacement. Equation 5.15, the mathematical formula for the hypergeometric distribution, gives the probability of X successes, given n, N and A.

HYPERGEOMETRIC DISTRIBUTION

$$P(X) = \frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}} \tag{5.15}$$

where $P(X)$ = the probability of X successes, given n, N and A
n = sample size
N = population size
A = number of successes in the population
$N - A$ = number of failures in the population
X = number of successes in the sample
$\binom{A}{X} = {}_AC_X$ (see Equation 5.10)

The number of successes in the sample, represented by X, cannot be greater than the number of successes in the population, A, or the sample size, n. Thus, the range of the hypergeometric random variable is limited to the minimum of the sample size or the number of successes in the population.

Equation 5.16 defines the mean of the hypergeometric distribution.

THE MEAN OF THE HYPERGEOMETRIC DISTRIBUTION

$$\mu = E(X) = \frac{nA}{N} \tag{5.16}$$

Equation 5.17 defines the standard deviation of the hypergeometric distribution.

THE STANDARD DEVIATION OF THE HYPERGEOMETRIC DISTRIBUTION

$$\sigma = \sqrt{\frac{nA(N-A)}{N^2}} \cdot \sqrt{\frac{N-n}{N-1}} \tag{5.17}$$

finite population correction factor
Factor required when sampling from a finite population without replacement.

In Equation 5.17, the expression $\sqrt{\frac{N-n}{N-1}}$ is a **finite population correction factor** that results from sampling without replacement from a finite population.

To illustrate the hypergeometric distribution, suppose that we wish to form a team of eight executives from different departments within a company. Suppose the company has a total of 30 executives, and 10 of these are from the finance department. If members of the team are to be selected at random, what is the probability that the team will contain two executives from the finance department? Here, the population of $N = 30$ executives within the company is finite. In addition, $A = 10$ are from the finance department and a team of $n = 8$ executives is to be selected.

Using Equation 5.15:

$$P(X=2)=\frac{\binom{10}{2}\binom{20}{6}}{\binom{30}{8}}=\frac{\frac{10!}{2!8!}\times\frac{20!}{6!14!}}{\frac{30!}{8!22!}}=0.2980\ldots$$

Thus, the probability that the team will contain two members from the finance department is 0.298, or 29.8%.

Such calculations can become tedious, especially as N gets larger. However, Microsoft Excel can be used to calculate hypergeometric probabilities. Figure 5.5, using the Excel 2010 and later inbuilt hypergeometric function HYPGEOM.DIST(sample_s,number_sample,population_s,number_population,cumulative), presents a Microsoft Excel worksheet for the team-formation example. Note that the number of executives from the finance department (i.e. the number of successes in the sample) can be equal to 0, 1, 2, … 8.

	A	B	
1	**Team formation analysis**		
2			
3	**Data**		
4	**Sample size**	**8**	
5	**No. of successes in population**	**10**	
6	**Population size**	**30**	
7			
8	**Hypergeometric probabilities table**		
9	**X**	**P(X)**	
10	**0**	**0.0215**	**=HYPEGEOM.DIST (A10, B4, B5, B6 FALSE)**
11	**1**	**0.1324**	**=HYPEGEOM.DIST (A11, B4, B5, B6 FALSE)**
12	**2**	**0.2980**	**=HYPEGEOM.DIST (A12, B4, B5, B6 FALSE)**
13	**3**	**0.3179**	**=HYPEGEOM.DIST (A13, B4, B5, B6 FALSE)**
14	**4**	**0.1738**	**=HYPEGEOM.DIST (A14, B4, B5, B6 FALSE)**
15	**5**	**0.0491**	**=HYPEGEOM.DIST (A15, B4, B5, B6 FALSE)**
16	**6**	**0.0068**	**=HYPEGEOM.DIST (A16, B4, B5, B6 FALSE)**
17	**7**	**0.0004**	**=HYPEGEOM.DIST (A17, B4, B5, B6 FALSE)**
18	**8**	**0.0000**	**=HYPEGEOM.DIST (A18, B4, B5, B6 FALSE)**

Figure 5.5
Microsoft Excel worksheet for the team-formation example

For earlier versions of Excel the corresponding hypergeometric function is HYPGEOMDIST (sample_s,number_sample,population_s,number_pop).

Problems for Section 5.5

LEARNING THE BASICS

5.33 Determine the following:

a. If $n = 4$, $N = 10$ and $A = 5$, find $P(X = 3)$.
b. If $n = 4$, $N = 6$ and $A = 3$, find $P(X = 1)$.
c. If $n = 5$, $N = 12$ and $A = 3$, find $P(X = 0)$.
d. If $n = 3$, $N = 10$ and $A = 3$, find $P(X = 3)$.

5.34 Referring to problem 5.33, calculate the mean and the standard deviation for the hypergeometric distributions described in (a) to (d).

APPLYING THE CONCEPTS

Problems 5.35 to 5.39 can be solved manually or by using Microsoft Excel.

5.35 An auditor for the Australian Taxation Office is selecting a sample of six tax returns from a batch of 100 for an audit. If two or more of these returns contain errors, the entire batch of 100 tax returns will be audited.

a. What is the probability that the entire batch will be audited if the true number of returns with errors in the batch is:
 i. 25?
 ii. 30?
 iii. 5?
 iv. 10?
b. Discuss the differences in your results depending on the true number of returns in the batch with error.

5.36 The dean of a business faculty wishes to form an executive committee of five from among the 40 tenured faculty members. The selection is to be random, and there are eight tenured faculty members in accounting.
a. What is the probability that the committee will contain:
 i. none of them?
 ii. at least one of them?
 iii. not more than one of them?
b. What is your answer to part (i) above if the committee consists of seven members?

5.37 In each game of OZ Lotto seven numbers are selected, from 1 to 45. Seven winning numbers are chosen at random plus two supplementary numbers. An extension of the hypergeometric distribution to calculate probabilities of selecting combinations of winning and supplementary numbers is:

$$P(X, Y) = \frac{\binom{A}{X}\binom{S}{Y}\binom{N-A-S}{n-X-Y}}{\binom{N}{n}}$$

where $P(X,Y)$ is the probability of selecting X winning numbers and Y supplementary numbers, and S is the number of supplementary numbers.
a. To win Division 1, the seven winning numbers must be selected. In any game, what is the probability of winning Division 1?
b. To win Division 2, six winning numbers plus either of the two supplementary numbers must be selected. In any game, what is the probability of winning Division 2?
c. To win Division 3, six winning numbers must be selected. In any game, what is the probability of winning Division 3?
d. To win Division 4, five winning numbers plus either of the two supplementary numbers must be selected. In any game, what is the probability of winning Division 4?
e. To win Division 5, five winning numbers must be selected. In any game, what is the probability of winning Division 5?
f. To win Division 6, four winning numbers must be selected. In any game, what is the probability of winning Division 6?
g. To win Division 7, three winning numbers plus either of the two supplementary numbers must be selected. In any game, what is the probability of winning Division 7?
h. What is the probability of selecting none of the winning or supplementary numbers?

5.38 In a certain game of Lotto six numbers are selected from 1 to 45. Six winning numbers are chosen at random plus two supplementary numbers. Use the formula in problem 5.37 or Equation 5.15 to calculate the following probabilities.
a. To win Division 1, the six winning numbers must be selected. In any game, what is the probability of winning Division 1?
b. To win Division 2, five winning numbers plus either of the two supplementary numbers must be selected. In any game, what is the probability of winning Division 2?
c. To win Division 3, five winning numbers must be selected. In any game, what is the probability of winning Division 3?
d. To win Division 4, four winning numbers must be selected. In any game, what is the probability of winning Division 4?
e. To win Division 5, three winning numbers plus either of the two supplementary numbers must be selected. In any game, what is the probability of winning Division 5?
f. What is the probability of selecting none of the winning or supplementary numbers?

5.39 In a shipment of 15 hard disks, five are defective. If four of the disks are inspected,
a. What is the probability that:
 i. exactly one is defective?
 ii. at least one is defective?
 iii. no more than two are defective?
b. What is the mean number of defective hard disks that you would expect to find in the sample of four hard disks?

5.6 (*ONLINE TOPIC*) USING THE POISSON DISTRIBUTION TO APPROXIMATE THE BINOMIAL DISTRIBUTION

Under certain circumstances the Poisson distribution can be used to approximate the binomial distribution.

Assess your progress 5

Summary

This chapter introduced mathematical expectation, covariance and the development and application of the binomial, Poisson and hypergeometric distributions. In the Gaia Cruises scenario, we saw how to calculate probabilities from the binomial and Poisson distributions concerning the number of enquiries converted to confirmed bookings in a sample of n enquiries and the number of online enquiries received in a given time interval. In the next chapter, important continuous distributions are introduced, in particular the normal distribution.

To help decide which discrete probability distribution to use for a particular situation, we need to ask the following questions:

- Is there a fixed number of observations n, each of which is classified as success or failure, or are we counting the number of times an event happens in an interval? If there is a fixed number of observations n, each of which is classified as success or failure, we can use the binomial or hypergeometric distribution, if the properties of the distribution are satisfied. If we are counting the number of events in an interval, we can use the Poisson distribution only if all its properties are satisfied.
- In deciding whether to use the binomial or hypergeometric distribution, is the probability of success constant for all observations? If yes, we may be able to use the binomial distribution. If no, we may be able to use the hypergeometric distribution.

Key formulas

Expected value μ of a discrete random variable

$$\mu = E(X) = \sum_{i=1}^{N} X_i P(X_i) \quad \textbf{(5.1)}$$

Variance of a discrete random variable

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i) \quad \textbf{(5.2a) (definition)}$$

$$\sigma^2 = \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2 \quad \textbf{(5.2b) (calculation)}$$

Standard deviation of a discrete random variable

$$\sigma = \sqrt{\sigma^2} \quad \textbf{(5.3)}$$

Covariance

$$\sigma_{XY} = \sum_{all\ X_i} \sum_{all\ Y_j} [X_i - E(X)][Y_j - E(Y)]P(X_i\ and\ Y_j) \quad \textbf{(5.4a) (definition)}$$

$$\sigma_{XY} = \sum_{all\ X_i} \sum_{all\ Y_j} X_i Y_j P(X_i\ and\ Y_j) - E(X)E(Y) \quad \textbf{(5.4b) (calculation)}$$

Expected value of the sum of two random variables

$$E(X + Y) = E(X) + E(Y) \quad \textbf{(5.5)}$$

Variance of the sum of two random variables

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} \quad \textbf{(5.6)}$$

Standard deviation of the sum of two random variables

$$\sigma_{X+Y} = \sqrt{\sigma^2_{X+Y}} \quad \textbf{(5.7)}$$

Portfolio expected return

$$E(P) = wE(X) + (1 - w)E(Y) \quad \textbf{(5.8)}$$

Portfolio risk

$$\sigma_p = \sqrt{w^2\sigma^2_X + (1 - w)^2\sigma^2_Y + 2w(1 - w)\sigma_{XY}} \quad \textbf{(5.9)}$$

Combinations

$$\binom{n}{X} = {}_nC_X = \frac{n!}{X!(n - X)!} \quad \textbf{(5.10)}$$

Binomial distribution

$$P(X) = \frac{n!}{X!(n - X)!} p^X(1 - p)^{n - X} \quad \textbf{(5.11)}$$

The mean of the binomial distribution

$$\mu = E(X) = np \quad \textbf{(5.12)}$$

The standard deviation of the binomial distribution

$$\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)} \quad \textbf{(5.13)}$$

Poisson distribution

$$P(X) = \frac{e^{-\lambda}\lambda^X}{X!} \quad \textbf{(5.14)}$$

Hypergeometric distribution

$$P(X) = \frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}} \quad \textbf{(5.15)}$$

The mean of the hypergeometric distribution

$$\mu = E(X) = \frac{nA}{N} \quad \textbf{(5.16)}$$

The standard deviation of the hypergeometric distribution

$$\sigma = \sqrt{\frac{nA(N-A)}{N^2}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \textbf{(5.17)}$$

Key terms

Chapter review problems

CHECKING YOUR UNDERSTANDING

5.40 What is the meaning of the expected value of a probability distribution?

5.41 What are the four properties of a binomial distribution?

5.42 What are the properties of a Poisson distribution?

5.43 When is the hypergeometric distribution used instead of the binomial distribution?

APPLYING THE CONCEPTS

Problems 5.44 to 5.53 can be solved manually or by using Microsoft Excel. Some, but not all, can also be solved using Tables E.6 and E.7.

5.44 From September 1984 to January 2014 the ASX All Ordinaries Index has opened higher than the previous month for 211 of the 353 months – that is, approximately 59.8% of months (YAHOO!7FINANCE <http://au.finance.yahoo.com> accessed January 2014).

a. Assuming a binomial distribution, estimate the probability that the ASX All Ordinaries Index will open higher than the previous month:

i. for one month

ii. for two months in a row

iii. in four of the next five months

iv. in none of the next five years

b. For the situation in (a) above, what assumption of the binomial distribution might not be valid?

5.45 At a recent election, 12% of the voters in a certain electorate gave their first preference to the Greens candidate. If 10 people on the electoral roll for that electorate were randomly selected, find the probability that:

a. exactly four gave their first preference to the Greens candidate

b. at most four gave their first preference to the Greens candidate

c. a majority gave their first preference to the Greens candidate

5.46 When calculating premiums on life insurance products, insurance companies often use life tables that enable the probability of a person dying in any age interval to be calculated.

The following data obtained from New Zealand Life Tables 2010–2012 gives the number out of 100,000 New Zealand-born males and females who are still alive during each five-year period of life between age 20 and 60 (inclusive).

Exact age (years)	Number alive at exact age: Out of 100,000 females born	Out of 100,000 males born
20	99,252	98,923
25	99,066	98,471
30	98,877	98,126
35	98,633	97,734
40	98,275	97,174
45	97,740	96,389
50	96,896	95,247
55	95,093	93,534
60	94,757	93,013

Data obtained from <www.stats.govt.nz> accessed January 2014. © Statistics New Zealand and licensed by Statistics New Zealand for re-use under the Creative Commons Attribution 3.0 New Zealand licence

Suppose a New Zealand-born female on her 30th birthday purchases a one million dollar, five-year term life policy from an insurance company. That is, the insurance company must pay her estate $1 million if she dies within the next five years.

a. Determine the insurance company's expected payout on this policy.
b. What would be the minimum you would expect the insurance company to charge her for this policy?
c. What would the expected payout be if the same policy were taken out by a New Zealand-born female on her 40th birthday?
d. Repeat parts (a) to (c) for a New Zealand-born male.

5.47 The emergency facility at a small country hospital has been in operation for 60 weeks and has been used 120 times. The weekly pattern of demand for this facility has a Poisson distribution. Find the:

a. mean demand per week
b. probability the emergency facility is not used in a given week
c. probability the emergency facility is used at least twice in a week
d. probability the room is used at least once in a given two-week period

5.48 Check$mart's records show that 58% of its customers pay only the minimum repayment on their credit card each month.

a. If a random sample of 20 credit-card holders is selected, what is the probability that:
  i. none pays the minimum amount?
  ii. no more than five pay the minimum amount?
  iii. more than 10 pay the minimum amount?
b. What assumptions did you have to make to answer each part of (a) above?

5.49 In 2012, 80% of New Zealand households had Internet access, with 75% of households having broadband connection (*Household Use of Information and Communication Technology: 2012,* Statistics New Zealand <www.stats.govt.nz> accessed January 2014).

Suppose 10 New Zealand households are randomly and independently sampled. Find the probability that:

a. fewer than nine households have home Internet access
b. at least eight households have a broadband connection

5.50 A customer service manager of Check$mart bank is monitoring one of its phone banking call centres servicing a rural region.

Suppose that on average the call centre receives 180 calls an hour during its operating hours of 8 am to 6 pm.

a. Can the Poisson distribution be used to model the number of calls received in one minute? Explain.
b. Assuming the number of calls received in a given interval is Poisson, calculate the probability that:
  i. in a given minute exactly two calls will be received
  ii. more than two calls will be received in a minute
  iii. the number of calls received in 5 minutes is at least 20
  iv. the number of calls received in 5 minutes is less than 10

At current staffing levels calls start to queue, increasing the time it takes to answer a call, when the number of calls received in 5 minutes is 20 or more. However, when there are less than 10 calls in 5 minutes, more than one Customer Service Officer is usually available, increasing unproductive staff time.

c. What conclusions can you draw from problem (b) parts (iii) and (iv) above?

5.51 Suppose the average number of students who log on to a university's computer system is 4.45 in each 5-minute interval.

a. What is the probability that six students will log on in the next minute?
b. What is the probability that fewer than six students will log on during the next two minutes?

5.52 A study of various news home pages reports that the mean number of bad links per home page is 0.4 and the mean number of spelling errors per home page is 0.16. Use the Poisson distribution to find the probability that a randomly selected home page will contain:

a. no bad links
b. five or more bad links
c. no spelling errors
d. 10 or more spelling errors

5.53 In an online test, 10 multiple-choice questions are randomly selected from a test bank of 100 questions.

Supposing that each student has two attempts at the online test, what is the probability that in the second test a student attempts there are:

a. no questions from the first test?
b. at least one question from the first test?
c. exactly five questions from the first test?
d. 10 questions from the first test?

5.54 The following table gives the grade distribution at a certain university.

Fail	Pass	Credit	Distinction	High Distinction
15%	40%	25%	15%	5%

Supposing that a result is selected randomly, what is the probability that:

a. the result is a passing grade (Pass or above)?
b. the result is a Credit or above?
c. If a random sample of 15 results is selected, what is the probability of:
  i. exactly three Fails?

ii. more than five Fails?
iii. all being Pass or above?
iv. none being Credit or above?
v. exactly five being Credits or above?
vi. at least one Distinction or High Distinction?

d. What is the expected number, variance and standard deviation of the number of:
i. Fail grades?
ii. grades Pass or above?

e. Comment on the relationship between (i) and (ii) in problem (d) above.

A grade point of 7 is assigned to each High Distinction, 6 to each Distinction, 5 to each Credit, 4 to each Pass and 0 to each fail.

f. What is the average, variance and standard deviation of grade points for the university?

5.55 You are trying to develop a strategy for investing in two different shares. The anticipated annual return for a $1,000 investment in each share has the following probability distribution:

	Returns	
Probability	**Share A**	**Share B**
0.25	$240	–$100
0.50	$150	$150
0.25	–$100	$240

a. Calculate:
i. the expected returns for share A and for share B
ii. the variances and standard deviations for share A and for share B
iii. the covariance of share A and share B

b. Suppose you want to create a portfolio that consists of share A and share B. Calculate the portfolio expected return and risk if the proportion invested in share A is:
i. 0.40
ii. 0.50
iii. 0.60

c. On the basis of the results in (b), which portfolio would you recommend? Explain.

5.56 The breakdown by home address of the previous year's 993 drink-driving offences in Problem 2.50 is:

Home address	Number of drink-driving offences
Local – in council area	
Seaside town	151
Not seaside town	462
Not local – not in council area	
Intrastate (within state)	130
Interstate (another state)	228
International (outside Australia)	22

Suppose that Kai selects at random 20 of the offenders to interview in depth. What is the probability that:

a. all 20 will be local?
b. 15 will be local?
c. five will be from interstate?
d. at least 10 will be not local?

Appendix 5

Using statistical software for discrete probability distributions

You can use either PHStat2 or the probability worksheets included with the data sets for the textbook to calculate the probabilities in this chapter.

Calculating expected value, variance and standard deviation of a discrete random variable

To calculate these parameters you must use the expected value worksheet, as neither Excel nor PHStat2 have commands for this calculation.

Using the Expected Value worksheet

Open the **Expected Value.xls** file. This worksheet already contains the entries for the Table 5.1 home mortgage approval example and uses the SUM and SQRT (square root) functions to calculate the parameters.

To adapt this worksheet to other problems:

- If you have more or less than seven outcomes, first add or delete table rows by selecting the cell range **A5:E5**, then right-click and select either **Insert** or **Delete**. If a box of options appears, choose **Shift cells down** if adding cells, or choose **Shift cells up** if deleting.

- If adding rows, copy the formulas in cell range **C4:E4** down through the new table rows.
- Enter a corrected list of *X* values in column A starting with cell A4.
- Enter the new values for *P*(*X*) in column B.

Calculating portfolio expected return and portfolio risk

You can either use the PHStat2 **Covariance and Portfolio Analysis** procedure or make entries in the portfolio worksheet, **Portfolio.xls**.

Using PHStat2 – covariance and portfolio analysis

Open the workbook in which you want your portfolio analysis worksheet to be placed.

Select **PHStat ➔ Decision-Making ➔ Covariance and Portfolio Analysis**. In the Covariance and Portfolio Management dialog box (see Figure A5.1) enter the **Number of Outcomes**, a **Title**, check **Portfolio Management Analysis**, and click on **OK**.

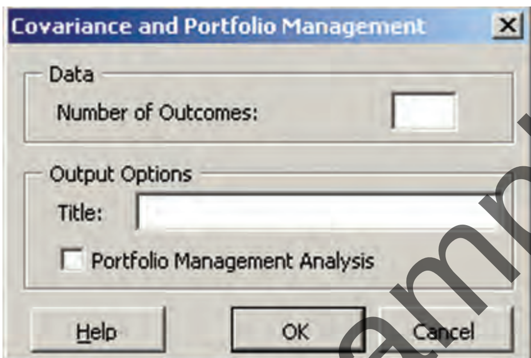

Figure A5.1 Covariance and Portfolio Management dialog box

In the worksheet created by PHStat2, enter in the tinted cells near the top the probabilities and outcome values in **Probabilities and Outcomes** and the weight value in **Weight Assigned to X**.

Using the Portfolio worksheet

Open the **Portfolio.xls** file. This worksheet already contains the entries for the investment data of Table 5.3. To adapt this worksheet to other problems:

- If you have more or less than three outcomes, first add or delete table rows by selecting **row 5**, then right-click and select either **Insert** or **Delete**. If a box of options appears, choose **Entire row**.
- If adding rows, copy the formulas in cell range **F4:J4** down through the new table rows.
- Enter the new probability and outcome values and the new **Weight Assigned to X** values in the tinted cells near the top of the worksheet.

Calculating binomial probabilities

You can use either the PHStat2 **Binomial** procedure or make entries in the binomial worksheet, **Binomial.xls**.

Using PHStat2 – Binomial

Open the workbook in which you want your binomial worksheet to be placed.

Select **PHStat ➔ Probability & Prob. Distributions ➔ Binomial**. In the Binomial Probability Distribution dialog box (see Figure A5.2) enter the **Sample Size**, the **Probability of Success** and the range of outcome required. Enter a **Title** and click on **OK**.

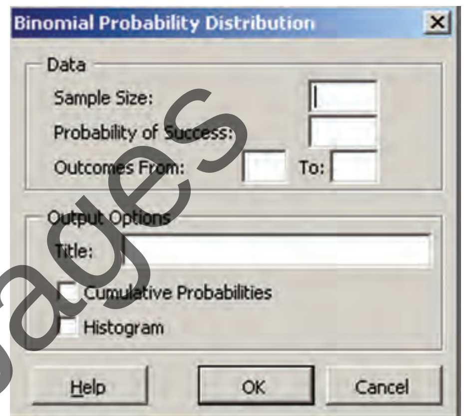

Figure A5.2 Binomial Probability Distribution dialog box

Using the Binomial worksheet

Open the **Binomial.xls** Excel file. This worksheet already contains the entries for the converted enquiries example used in Section 5.3. The worksheet Binomial Excel 2007 uses the Excel 2007 function BINOMDIST to calculate binomial probabilities. In the worksheet Binomial Excel 2010 BINOM.DIST is used to calculate these probabilities.

To adapt this worksheet to other problems:

- If your sample size is other than 4, first add or delete table rows by selecting **row 15**, then right-click and select either **Insert** or **Delete**. If a box of options appears, choose **Entire row**.
- Copy the entries in cell range **A14:F14** down through the entire table to update the table.
- Enter the new sample size and probability of success values in cells B4 and B5.

Calculating Poisson probabilities

You can either use the PHStat2 **Poisson** procedure or make entries in the Poisson worksheet, **Poisson.xls**.

Using PHStat2 – Poisson

Open the workbook in which you want the Poisson worksheet to be placed.

Select **PHStat ➔ Probability & Prob. Distributions ➔ Poisson**. In the Poisson Probability Distribution dialog box (see Figure A5.3) enter the **Average/Expected number of Sucesses**, a **Title**, and click on **OK**.

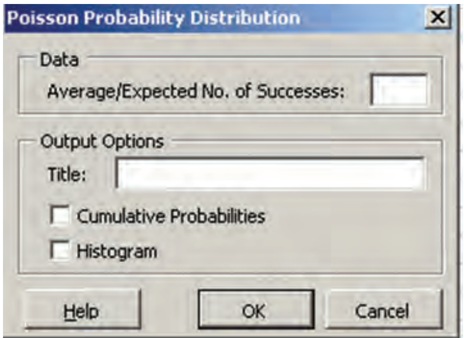

Figure A5.3 Poisson Probability Distribution dialog box

Using the Poisson worksheet

Open the **Poisson.xls** Excel file. This worksheet already contains the entries for the online enquiries received example used in Section 5.4. The worksheet Poisson Excel 2007 uses the Excel 2007 function POISSON to calculate Poisson probabilities. In the worksheet Poisson Excel 2010 POISSON.DIST is used to calculate these probabilities.

To adapt this worksheet to other problems:

- If you need more than 20 occurrences, first add table rows by selecting **row 9**, then right-click and select **Insert**. If a box of options appears, choose **Entire row**.
- Copy the entries in cell range **A8:F8** down through the entire table to update the table.
- Enter the new **Average/Expected number of successes** value in cell E4.

Calculating hypergeometric probabilities

You can either use the PHStat2 **Hypergeometric** procedure or make entries in the hypergeometric worksheet, **Hypergeometric.xls**.

Using PHStat2 – hypergeometric

Open the workbook in which you want the Hypergeometric worksheet to be placed.

Select **PHStat ➔ Probability & Prob. Distributions ➔ Hypergeometric**. In the Hypergeometric Probability Distribution dialog box (see Figure A5.4) enter the **Sample Size**, the **No. of Successes in Population**, and **Population Size**. Enter a **Title** and click on **OK**.

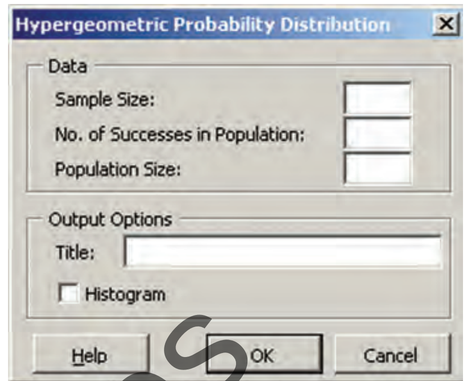

Figure A5.4 Hypergeometric Probability Distribution dialog box

Using the Hypergeometric worksheet

Open the **Hypergeometric.xls** Excel file. This worksheet already contains the entries for the team-formation example used in Section 5.5. The worksheet Hypergeometric Excel 2007 uses the Excel 2007 function HYPGEOMDIST to calculate hypergeometric probabilities. In the worksheet Hypergeometric Excel 2010 HYPGEOM.DIST is used to calculate these probabilities.

To adapt this worksheet to other problems:

- If your sample size is more or less than 8, first add or delete table rows by selecting **row 11**, then right-click and select either **Insert** or **Delete**. If a box of options appears, choose **Entire row**.
- Copy the entries in cell range **A10:C10** down through the entire table to update the table.
- Enter the new values for **Sample size**, **No. of successes in population** and **Population size** in the cell range **B4:B6**.