Endorsed for
Pearson Edexcel
Qualifications

MATHEMATICS

eBook
included

PEARSON EDEXCEL INTERNATIONAL A LEVEL

STATISTICS 2

STUDENT BOOK

PEARSON EDEXCEL INTERNATIONAL A LEVEL

STATISTICS 2

Student Book

Series Editors: Joe Skrakowski and Harry Smith

Authors: Greg Attwood, Tom Begley, Ian Bettison, Alan Clegg, Martin Crozier, Gill Dyer, Jane Dyer, Keith Gallick, Susan Hooker, Michael Jennings, John Kinoulty, Mohammed Ladak, Guilherme Frederico Lima, Jean Littlewood, Bronwen Moran, James Nicholson, Su Nicholson, Laurence Pateman, Keith Pledger, Joe Skrakowski, Harry Smith

Published by Pearson Education Limited, 80 Strand, London, WC2R 0RL.

www.pearsonglobalschools.com

Copies of official specifications for all Pearson qualifications may be found on the website: https://qualifications.pearson.com

Edited by Eric Pradel
Typeset by Tech-Set Ltd, Gateshead, UK

First published 2019

22 21 20 19
10 9 8 7 6 5 4 3 2 1

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 978 1 292245 17 1

Printed in Slovakia by Neografia

Picture Credits
The authors and publisher would like to thank the following individuals and organisations for permission to reproduce photographs:

Shutterstock.com: Pavlovska Yevheniia 1, katjen 35, TK Kurikawa 87, Sergey Mironov 102, Shaiith 112; **Science Photo Library:** ALAN DYER/VWPICS 17; **123RF:** Viparat Kluengsuwanchai 49

Cover images: *Front*: **Getty Images:** Werner Van Steen
Inside front cover: **Shutterstock.com:** Dmitry Lobanov

Endorsement Statement
In order to ensure that this resource offers high-quality support for the associated Pearson qualification, it has been through a review process by the awarding body. This process confirms that this resource fully covers the teaching and learning content of the specification or part of a specification at which it is aimed. It also confirms that it demonstrates an appropriate balance between the development of subject skills, knowledge and understanding, in addition to preparation for assessment.

Endorsement does not cover any guidance on assessment activities or processes (e.g. practice questions or advice on how to answer assessment questions) included in the resource, nor does it prescribe any particular approach to the teaching or delivery of a related course.

While the publishers have made every attempt to ensure that advice on the qualification and its assessment is accurate, the official specification and associated assessment guidance materials are the only authoritative source of information and should always be referred to for definitive guidance.

Pearson examiners have not contributed to any sections in this resource relevant to examination papers for which they have responsibility.

Examiners will not use endorsed resources as a source of material for any assessment set by Pearson. Endorsement of a resource does not mean that the resource is required to achieve this Pearson qualification, nor does it mean that it is the only suitable material available to support the qualification, and any resource lists produced by the awarding body shall include this and other appropriate resources.

ABOUT THIS BOOK

The following three themes have been fully integrated throughout the Pearson Edexcel International Advanced Level in Mathematics series, so they can be applied alongside your learning.

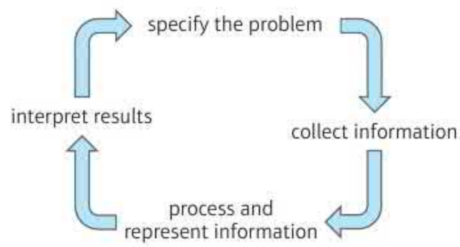
1. Mathematical argument, language and proof

- Rigorous and consistent approach throughout
- Notation boxes explain key mathematical language and symbols

2. Mathematical problem-solving

- Hundreds of problem-solving questions, fully integrated into the main exercises
- Problem-solving boxes provide tips and strategies
- Challenge questions provide extra stretch

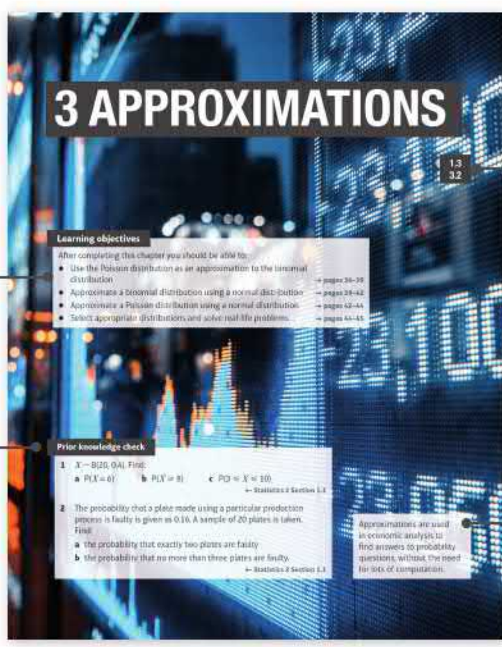
The Mathematical Problem-Solving Cycle



3. Transferable skills

- Transferable skills are embedded throughout this book, in the exercises and in some examples
- These skills are signposted to show students which skills they are using and developing

Finding your way around the book



Each chapter starts with a list of *Learning objectives*

The *Prior knowledge check* helps make sure you are ready to start the chapter

Glossary terms will be identified by bold blue text on their first appearance

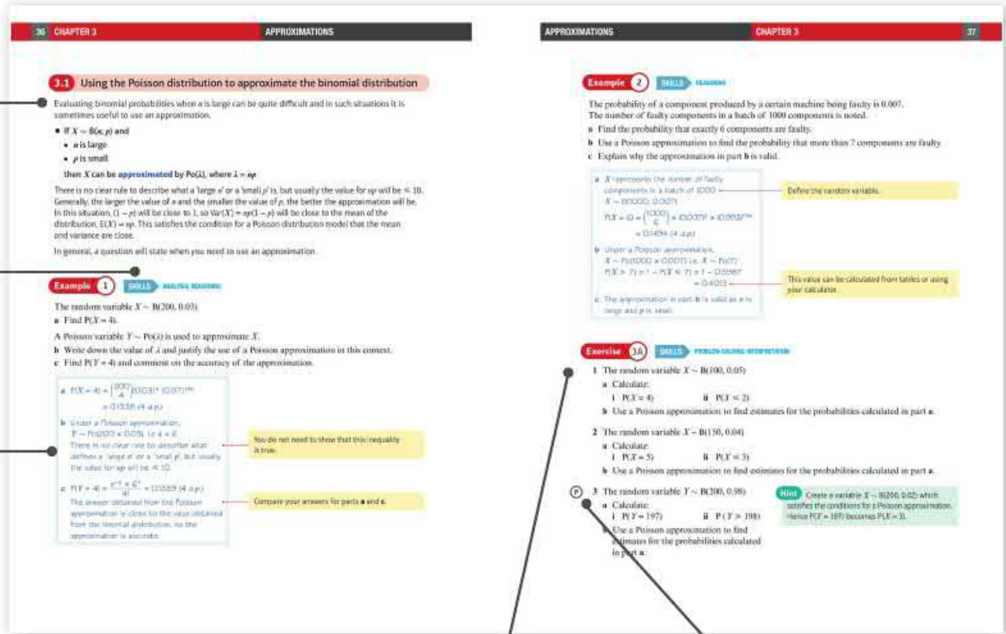
Each chapter is mapped to the specification content for easy reference

The real world applications of the maths you are about to learn are highlighted at the start of the chapter

Each section begins with an explanation and key learning points

Transferable skills are signposted where they naturally occur in the exercises and examples

Step-by-step worked examples focus on the key types of questions you'll need to tackle



Problem-solving boxes provide hints, tips and strategies, and *Watch out* boxes highlight areas where students often lose marks in their exams

Exercises are packed with exam-style questions to ensure you are ready for the exams

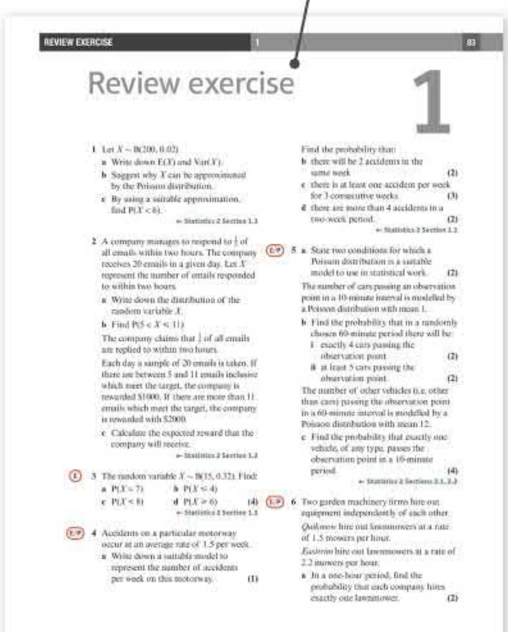
Exercise questions are carefully graded so they increase in difficulty and gradually bring you up to exam standard

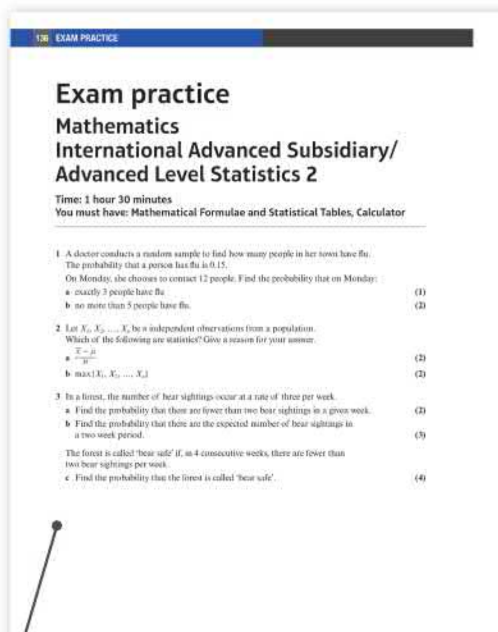
Exam-style questions are flagged with Ⓔ

Problem-solving questions are flagged with Ⓟ

Each chapter ends with a *Chapter review* and a *Summary of key points*

After every few chapters, a *Review exercise* helps you consolidate your learning with lots of exam-style questions





A full practice paper at the back of the book helps you prepare for the real thing

Relationship of assessment objectives to units

	Assessment objective				
S2	AO1	AO2	AO3	AO4	AO5
Marks out of 75	25–30	20–25	10–15	5–10	5–10
%	$33\frac{1}{3}$–40	$26\frac{2}{3}$–$33\frac{1}{3}$	$13\frac{1}{3}$–20	$6\frac{2}{3}$–$13\frac{1}{3}$	$6\frac{2}{3}$–$13\frac{1}{3}$

Calculators

Students may use a calculator in assessments for these qualifications. Centres are responsible for making sure that calculators used by their students meet the requirements given in the table below.

Students are expected to have available a calculator with at least the following keys: +, –, ×, ÷, π, x^2, $\sqrt{x}$, $\frac{1}{x}$, x^y, ln x, e^x, $x!$, sine, cosine and tangent and their inverses in degrees and decimals of a degree, and in radians; memory.

Prohibitions

Calculators with any of the following facilities are prohibited in all examinations:

- databanks
- retrieval of text or formulae
- built-in symbolic algebra manipulations
- symbolic differentiation and/or integration
- language translators
- communication with other machines or the internet

Extra online content

Whenever you see an *Online* box, it means that there is extra online content available to support you.

SolutionBank

SolutionBank provides worked solutions for questions in the book. Download the solutions as a PDF or quickly find the solution you need online.

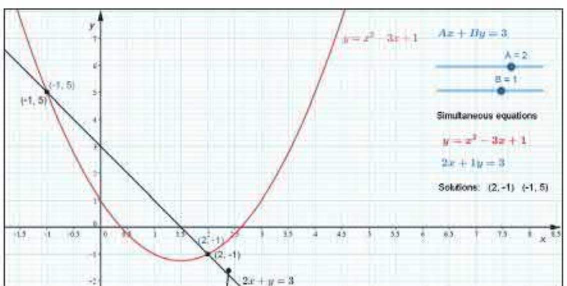
Use of technology

Explore topics in more detail, visualise problems and consolidate your understanding. Use pre-made GeoGebra activities or Casio resources for a graphic calculator.

Online Find the point of intersection graphically using technology.

GeoGebra

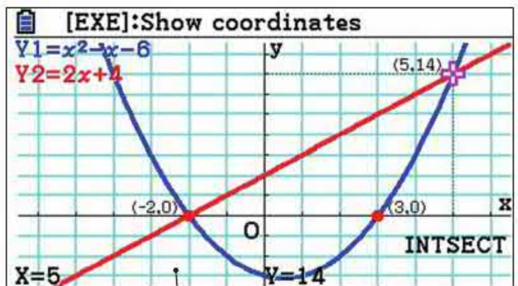
GeoGebra-powered interactives



Interact with the maths you are learning using GeoGebra's easy-to-use tools

CASIO

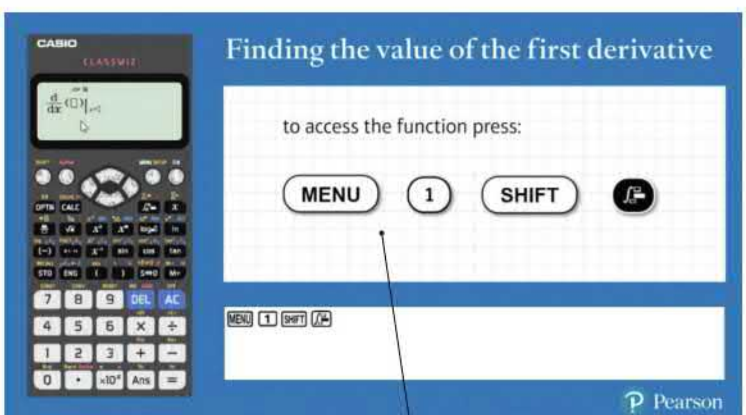
Graphic calculator interactives



Explore the maths you are learning and gain confidence in using a graphic calculator

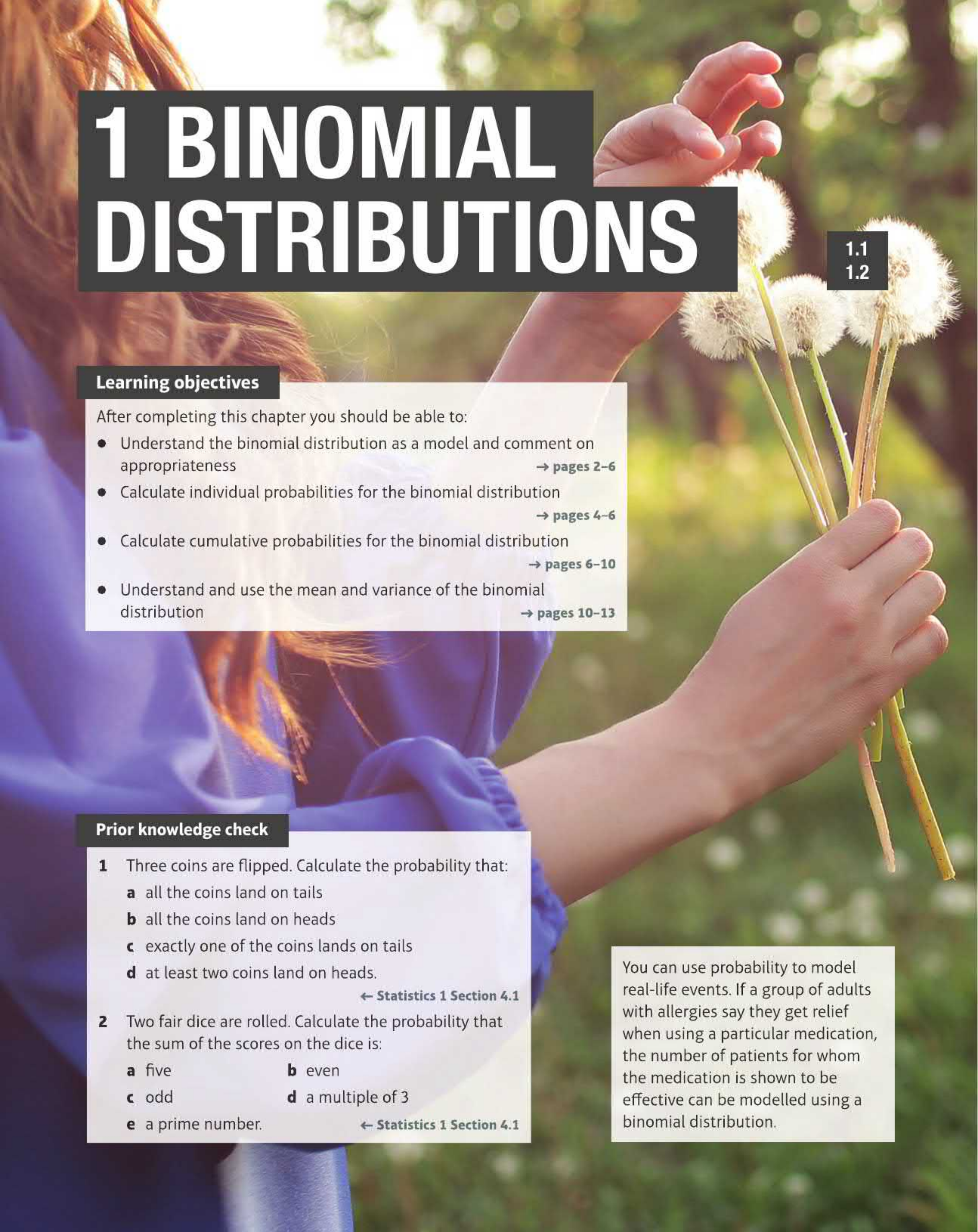
Calculator tutorials

Our helpful video tutorials will guide you through how to use your calculator in the exams. They cover both Casio's scientific and colour graphic calculators.



Step-by-step guide with audio instructions on exactly which buttons to press and what should appear on your calculator's screen

Online Work out each coefficient quickly using the ^nC_r and power functions on your calculator.



1 BINOMIAL DISTRIBUTIONS

1.1
1.2

Learning objectives

After completing this chapter you should be able to:

- Understand the binomial distribution as a model and comment on appropriateness → pages 2–6
- Calculate individual probabilities for the binomial distribution → pages 4–6
- Calculate cumulative probabilities for the binomial distribution → pages 6–10
- Understand and use the mean and variance of the binomial distribution → pages 10–13

Prior knowledge check

1 Three coins are flipped. Calculate the probability that:

a all the coins land on tails

b all the coins land on heads

c exactly one of the coins lands on tails

d at least two coins land on heads.

← Statistics 1 Section 4.1

2 Two fair dice are rolled. Calculate the probability that the sum of the scores on the dice is:

a five **b** even

c odd **d** a multiple of 3

e a prime number.

← Statistics 1 Section 4.1

You can use probability to model real-life events. If a group of adults with allergies say they get relief when using a particular medication, the number of patients for whom the medication is shown to be effective can be modelled using a binomial distribution.

1.1 The binomial distribution

Consider a biased dice whose **probability** of rolling a 6 is equal to p.

You are going to roll the dice 4 times and count how many times you roll a 6.

The number of 6s that you roll is a **discrete random variable**. We can define it as:

Let X represent the number of 6s rolled when we roll a dice 4 times.

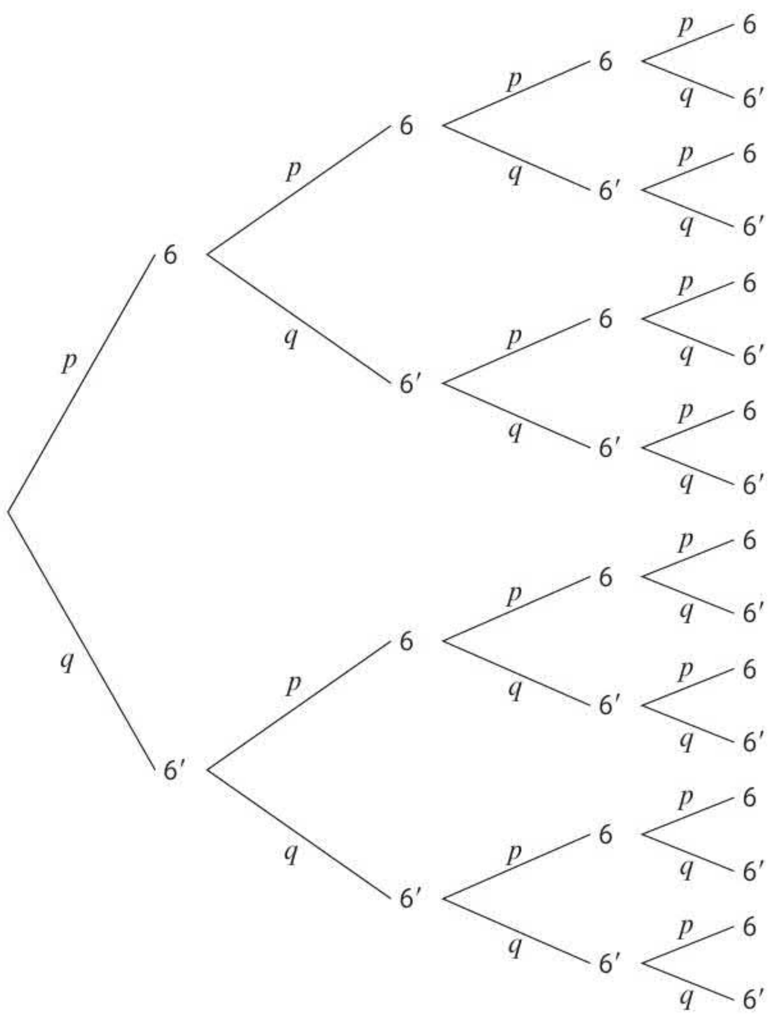
There are a few observations we can make:

- There are a fixed number of **trials** – we roll the dice 4 times
- There are two **outcomes** – rolling a 6 and not rolling a 6
- The trials are **independent** – each trial has no effect on future trials
- The probability of rolling a 6 is **constant**.

Situations like this can be modelled using a tree diagram to represent the possible outcomes.

Let the probability of rolling a 6 $= p$

Let the probability of rolling (not a six) $= q$, where $p + q = 1$



Number of 6s	Probability
4	p^4
3	p^3q
3	p^3q
2	p^2q^2
3	p^3q
2	p^2q^2
2	p^2q^2
1	pq^3
3	p^3q
2	p^2q^2
2	p^2q^2
1	pq^3
2	p^2q^2
1	pq^3
1	pq^3
0	q^4

We can write this as a table.

x	0	1	2	3	4
$\mathrm{P}(X = x)$	q^4	$4q^3p$	$6p^2q^2$	$4p^3q$	p^4

Consider the expansion of :

$$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

You will notice that each term in the sum corresponds to the probabilities in the distribution.

For example, $\mathrm{P}(X = 3) = 4p^3q$

$$(p + q)^4 = \binom{4}{4}p^4q^0 + \binom{4}{3}p^3q^1 + \binom{4}{2}p^2q^2 + \binom{4}{1}p^1q^3 + \binom{4}{0}p^0q^4$$

We can create a general formula for this case:

$$\mathrm{P}(X = x) = \binom{4}{x}p^xq^{4-x}$$

If we were to roll the dice n times, the expansion would be:

$$(p + q)^n = \binom{n}{n}p^n + \binom{n}{n-1}p^{n-1}q + \binom{n}{n-2}p^{n-2}q^2 + \ldots + \binom{n}{2}p^2q^{n-2} + \binom{n}{1}pq^{n-1} + \binom{n}{n}p^n$$

Each probability can be written as:

$$\mathrm{P}(X = x) = \binom{n}{x}p^nq^{n-x}$$

Since $p + q = 1$, the sum of all of these probabilities is 1, as required.

Setting $q = 1 - p$, we can also write the probabilities as:

$$\mathrm{P}(X = x) = \binom{n}{x}p^n(1 - p)^{n-x}$$

When you carry out a number of trials in an experiment or survey, you can define a **random variable** X to represent the number of **successful** trials.

- You can **model** X with a **binomial distribution**, $\mathrm{B}(n, p)$, if:
  - there are a fixed number of trials, n
  - there are two possible outcomes (success and failure)
  - there is a fixed **probability** of success, p
  - the trials are independent of each other

Hint If P(Success) $= p$ and there are only two outcomes then P(Failure) $= 1 - p = q$

- If a random variable X has the binomial distribution $\mathrm{B}(n, p)$ then its probability is given by:

$$\mathrm{P}(X = r) = \binom{n}{r}p^r(1 - p)^{n-r}$$

Notation You write $X \sim \mathrm{B}(n, p)$

n is sometimes called the **index** and p is sometimes called the **parameter**.

Links $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

It is sometimes written as nCr or ${}^nC_r$. It represents the number of ways of selecting r successful outcomes from n trials.

← Pure 2 Section 4.2

You can use your calculator to work out binomial probabilities. You can either use the rule given above, together with the ${}^nC_r$ function, or use the binomial probability distribution function directly.

Example 1 SKILLS INTERPRETATION

The random variable $X \sim B\left(12, \frac{1}{6}\right)$. Find:

a $P(X = 2)$ **b** $P(X = 9)$ **c** $P(X \leqslant 1)$

a $P(X = 2) = \binom{12}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{10} = \frac{12!}{2!10!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{10}$

$= 0.296\,09\ldots$

$= 0.296$ (3 significant figures)

Use the formula with $n = 12$, $p = \frac{1}{6}$ and $x = 2$.

b $P(X = 9) = \binom{12}{9}\left(\frac{1}{6}\right)^9\left(\frac{5}{6}\right)^3$

$= 0.000\,012\,63\ldots$

$= 0.000\,012\,6$ (3 s.f.)

Use the formula with $n = 12$, $p = \frac{1}{6}$ and $x = 9$.

c $P(X \leqslant 1) = P(X = 0) + P(X = 1)$

$= \left(\frac{5}{6}\right)^{12} + \binom{12}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{11}$

$= 0.112\,156\ldots + 0.269\,17\ldots$

$= 0.381\,33\ldots$

$= 0.381$ (3 s.f.)

A binomial distribution can take any value from 0 up to n **inclusive**. So there are two possible outcomes that **satisfy** the inequality: $X = 0$ and $X = 1$.

Online Use the ${}^nC_r$ function on your calculator to work out binomial probabilities.

Example 2 SKILLS ANALYSIS

The probability that a randomly chosen member of a reading group is left-handed is 0.15. A random **sample** of 20 members of the group is taken.

a Suggest a suitable model for the random variable X, the number of members in the sample who are left-handed. Justify your choice.

b Use your model to calculate the probability that:

i exactly 7 of the members in the sample are left-handed

ii fewer than two of the members in the sample are left-handed.

a The random variable can take two values, left-handed or right-handed.
There are a fixed number of trials, 20, and a fixed probability of success, 0.15.
Assuming each member in the sample is independent, a suitable model is
$X \sim B(20, 0.15)$

A binomial model is a suitable choice. State the **assumptions** that are necessary for the binomial model, and make sure that you specify the values of n and p.

b i $P(X = 7) = \binom{20}{7} \times (0.15)^7(0.85)^{13}$

$= 0.016\,01\ldots$

$= 0.0160$ (3 s.f.)

Online Work this out directly using the binomial probability distribution function on your calculator and entering $x = 7$, $n = 20$ and $p = 0.15$.

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

$= 0.038\,75\ldots + 0.136\,79\ldots$

$= 0.176$ (3 s.f.)

In this situation, 'fewer than two' means 0 or 1.

Exercise 1A SKILLS PROBLEM-SOLVING

1 The random variable $X \sim \mathrm{B}\left(8, \frac{1}{3}\right)$. Find:

a $\mathrm{P}(X = 2)$ **b** $\mathrm{P}(X = 5)$ **c** $\mathrm{P}(X \leqslant 1)$

2 The random variable $T \sim \mathrm{B}\left(15, \frac{2}{3}\right)$. Find:

a $\mathrm{P}(T = 5)$ **b** $\mathrm{P}(T = 10)$ **c** $\mathrm{P}(3 \leqslant T \leqslant 4)$

3 A student suggests using a binomial distribution to model the following situations. Define the random variable, state any assumptions that must be made and give possible values for n and p.

a A sample of 20 bolts from a large batch (group) is checked for faults. It is estimated that 1% of all bolts produced will be faulty.

b Some traffic lights have three settings: stop (red) 48% of the time, wait or get ready (orange) 4% of the time, and go (green) 48% of the time. Assuming that you proceed forward only when the light is green, model the number of times that you have to wait or stop on a journey passing through 6 sets of traffic lights.

c When Stephanie plays tennis with Rashmi, one in eight of her serves, on average, is an 'ace'. Stephanie serves 30 times. How many aces does she serve?

4 State which of the following can be modelled with a binomial distribution and which cannot. Give reasons for your answers.

a Given that 15% of people have blood that is Rhesus negative (Rh^-), model the number of students in a statistics class of 14 who are Rh^-.

b You are given a fair coin and told to keep flipping it until you obtain 4 heads in a row. Model the number of flips you need.

c A certain car manufacturer produces 12% of new cars in the colour red, 8% in blue, 15% in white and the rest in other colours. You make a note of the colour of the first 15 new cars of this make. Model the number of red cars you observe.

(P) **5** A balloon manufacturer claims that 95% of his balloons will not burst when blown up. You have 20 of these balloons to blow up for a birthday party.

a What is the probability that none of them burst when blown up?

b Find the probability that exactly 2 balloons burst.

(E/P) **6** The probability of a switch being faulty is 0.08. A random sample of 10 switches is taken from the production line.

a Define a suitable distribution to model the number of faulty switches in this sample, and justify your choice. **(2 marks)**

b Find the probability that the sample contains 4 faulty switches. **(2 marks)**

E/P **7** A particular genetic marker is present in 4% of the population.

a State any assumptions that are required to model the number of people with this genetic marker in a sample of size n as a binomial distribution. **(2 marks)**

b Using this model, find the probability of exactly 6 people having this marker in a sample of size 50. **(2 marks)**

E/P **8** A dice is biased so that the probability of it landing on a six is 0.3. Vivek rolls the dice 15 times.

a State any assumptions that are required to model the number of sixes as a binomial distribution. State the distribution. **(2 marks)**

b Find the probability that Vivek rolls exactly 4 sixes. **(2 marks)**

c Find the probability that he rolls 2 or fewer sixes. **(3 marks)**

1.2 Cumulative probabilities

A **cumulative probability function** for a random variable X tells you the sum of all the individual probabilities up to and including the given value of x in the calculation for $P(X \leqslant x)$.

For the binomial distribution $X \sim B(n, p)$ there are tables in the formula book giving $P(X \leqslant x)$ for various values of n and p. An extract from the tables is shown below:

$p =$	0.05	0.10	0.15	0.20	0.25	0.30
$n = 5, x = 0$	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681
1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282
2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369
3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692
4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976
$n = 6, x = 0$	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176

For the binomial distribution $X \sim B(5, 0.3)$, this tells you that $P(X \leqslant 2) = 0.8369$ (4 d.p.)

You might have a calculator that has a binomial cumulative function for any values of x, n and p.

Example 3 SKILLS INTERPRETATION

The random variable $X \sim B(20, 0.4)$. Find:

a $P(X \leqslant 7)$ **b** $P(X < 6)$ **c** $P(X \geqslant 15)$

a $P(X \leqslant 7) = 0.4159$

Use $n = 20$, $p = 0.4$ and $x = 7$. You can use tables or your calculator.

b $P(X < 6) = P(X \leqslant 5)$
$= 0.1256$

X can only take whole number values, so $P(X < 6) = P(X \leqslant 5)$.

c $P(X \geqslant 15) = 1 - P(X \leqslant 14)$
$= 1 - 0.9984$
$= 0.0016$

Hint When necessary, write your probabilities in the form $X \leqslant \ldots$ to make sure that you use the tables correctly.

When questions are set in context, there are different phrases that can be used to ask for probabilities. The correct understanding of these phrases is critical, especially when dealing with **cumulative** probabilities. The table below gives some examples.

Phrase	Means	Calculation
… greater than 5 …	$X > 5$	$1 - \mathrm{P}(X \leqslant 5)$
… no more than 3 …	$X \leqslant 3$	$\mathrm{P}(X \leqslant 3)$
… at least 7 …	$X \geqslant 7$	$1 - \mathrm{P}(X \leqslant 6)$
… fewer than 10 …	$X < 10$	$\mathrm{P}(X \leqslant 9)$
… at most 8 …	$X \leqslant 8$	$\mathrm{P}(X \leqslant 8)$

Hint It is very important to know these phrases and how to interpret and calculate them.

Example 4

SKILLS PROBLEM-SOLVING

A spinner is designed such that the probability it lands on red is 0.3. Sakiya has 12 spins. Find the probability that Sakiya obtains:

a no more than 2 reds

b at least 5 reds.

Sakiya decides to use this spinner for a class competition. She wants the probability of winning a prize to be less than 0.05. Each class member will have 12 spins and the number of reds will be recorded.

c Find how many reds are needed to win a prize.

Let X represent the number of reds in 12 spins.

$X \sim \mathrm{B}(12, 0.3)$

a $\mathrm{P}(X \leqslant 2) = 0.2528$

'no more than 2' means $X \leqslant 2$

b $\mathrm{P}(X \geqslant 5) = 1 - \mathrm{P}(X \leqslant 4)$

$= 1 - 0.7237$

$= 0.2763$

'at least 5' means $X \geqslant 5$

c Let r represent the smallest number of reds needed to win a prize.

Require: $\mathrm{P}(X \geqslant r) < 0.05$

Form a probability statement to represent the **condition** for winning a prize.

From tables:

$\mathrm{P}(X \leqslant 5) = 0.8822$

$\mathrm{P}(X \leqslant 6) = 0.9614$

$\mathrm{P}(X \leqslant 7) = 0.9905$

Problem-solving

When you are looking for the first probability that is greater or less than a given **threshold**, it is usually quicker to look on the tables rather than calculate them.

So: $\mathrm{P}(X \leqslant 6) = 0.9614$ implies that

$\mathrm{P}(X \geqslant 7) = 1 - 0.9614$

$= 0.0386 < 0.05$

Since $x = 6$ gives the first value greater than 0.95, use this probability and find $r = 7$.

So 7 or more reds will win a prize.

Always make sure that your final answer is related back to the context of the original question.

Sometimes it is useful to redefine the random variable so that the tables can be used.

Example 5 SKILLS ADAPTIVE LEARNING

Consider the discrete random variable $X \sim B(10, 0.7)$

Find:

a $P(X = 4)$

b $P(X \leqslant 6)$

a $P(X = 4) = {}^{10}C_4(0.7)^4 \times (0.3)^{10-4}$

$= 0.036\,756\,909$

$= 0.0368$ (3 s.f.)

Use the formula to calculate the probability.

b For this question, the table cannot be used with X since the values for the parameter p in the table are below 0.5. We can therefore redefine the random variable and adjust the question.

Let $Y \sim B(10, 0.3)$

Consider changing the 'success' over to give a parameter that is in the tables.

X:	0	1	2	3	4	5	6	7	8	9	10
Y:	10	9	8	7	6	5	4	3	2	1	0

Be careful to make sure you are finding the same probability.

And so $P(X \leqslant 6) = P(Y \geqslant 4)$

$= 1 - P(Y \leqslant 3)$

$= 1 - 0.6496$

$= 0.350$ (3 s.f.)

Turn it into a cumulative probability and look up from the table.

We can model a situation using the binomial distribution by making use of the assumptions of the binomial.

Consider selecting sweets from a bag without replacing what is taken. The bag contains a large number of strawberry-flavoured sweets and a large number of apple-flavoured sweets. We are also told that the sweets are in the ratio 3 : 1.

The probabilities will be affected each time a sweet is taken because they are not being replaced. However, the large number of sweets means that the probabilities will be small enough to be able to assume that the probability of success is constant.

Example 6 SKILLS INTERPRETATION

A bag contains a large number of strawberry-flavoured and apple-flavoured sweets in the ratio 1 : 4.

5 sweets are chosen at random without replacement.

a What assumptions need to be made to use the binomial distribution?

b With this assumption, find the probability that at least 4 of the sweets chosen are strawberry-flavoured.

a Since we are told that there are a large number of sweets, we can make the assumption that the probability of choosing a strawberry-flavoured sweet is constant.

Here, 'large' means large enough to make the assumption.

b Let X be the number of strawberry-flavoured sweets from the 5 sweets chosen.

Define the random variable.

Then

$X \sim \text{B}(5, 0.2)$

The ratio 1 : 4 means the probability of choosing a strawberry-flavoured sweet is $\frac{1}{1+4} = 0.2$

$\text{P}(X \geqslant 4)$

Write the question in terms of the random variable.

$\text{P}(X \geqslant 4) = 1 - \text{P}(X \leqslant 3)$

$= 1 - 0.9933$

Using tables

$= 0.0067$

Exercise 1B

SKILLS PROBLEM-SOLVING; INTERPRETATION

1 The random variable $X \sim \text{B}(9, 0.2)$. Find:

a $\text{P}(X \leqslant 4)$ **b** $\text{P}(X < 3)$ **c** $\text{P}(X \geqslant 2)$ **d** $\text{P}(X = 1)$

2 The random variable $X \sim \text{B}(20, 0.35)$. Find:

a $\text{P}(X \leqslant 10)$ **b** $\text{P}(X > 6)$

c $\text{P}(X = 5)$ **d** $\text{P}(2 \leqslant X \leqslant 7)$

Hint $\text{P}(2 \leqslant X \leqslant 7) = \text{P}(X \leqslant 7) - \text{P}(X \leqslant 1)$

3 The random variable $X \sim \text{B}(40, 0.4)$. Find:

a $\text{P}(X < 20)$ **b** $\text{P}(X > 16)$ **c** $\text{P}(11 \leqslant X \leqslant 15)$ **d** $\text{P}(10 < X < 17)$

4 The random variable $X \sim \text{B}(37, 0.85)$. Find:

a $\text{P}(X \leqslant 35)$ **b** $\text{P}(X = 23)$ **c** $\text{P}(25 < X \leqslant 27)$

(P) **5** Eight fair coins are tossed and the total number of heads showing is recorded. Find the probability of getting:

a no heads **b** at least 2 heads **c** more heads than tails.

(P) **6** For a particular type of plant, 25% have blue flowers. A garden centre sells these plants in trays of 15 plants of mixed colours. A tray is selected at random. Find the probability that the number of plants with blue flowers in this tray is:

a exactly 4

b at most 3

c between 3 and 6 (inclusive).

E/P **7** The random variable $X \sim B(50, 0.40)$. Find:

a the largest value of k such that $P(X \leqslant k) < 0.05$ **(1 mark)**

b the smallest number r such that $P(X > r) < 0.01$. **(2 marks)**

E/P **8** The random variable $X \sim B(40, 0.10)$. Find:

a the largest value of k such that $P(X < k) < 0.02$ **(1 mark)**

b the smallest number r such that $P(X > r) < 0.01$ **(2 marks)**

c $P(k \leqslant X \leqslant r)$. **(2 marks)**

E/P **9** In a town, 30% of residents listen to local radio.
Ten residents are chosen at random.
Let X represent the number of these 10 residents that listen to local radio.

a Suggest a suitable distribution for X and comment on any necessary assumptions. **(2 marks)**

b Find the probability that at least half of these 10 residents listen to local radio. **(2 marks)**

c Find the smallest value of s so that $P(X \geqslant s) < 0.01$ **(2 marks)**

E/P **10** A factory produces a **component** for the motor trade and 5% of the components are faulty. An employee regularly checks a random sample of 50 components. Find the probability that the next sample contains:

a fewer than 2 faulty components **(1 mark)**

b more than 5 faulty components. **(2 marks)**

The employee will stop production if the number of faulty components in the sample is greater than a certain value d. Given that the employee stops production less than 5% of the time:

c find the smallest value of d. **(2 marks)**

1.3 Mean and variance of the binomial distribution

When we use a distribution, it is very useful to be able to discuss a measure of centrality and a measure of dispersion.

Links For a probability distribution, these were the expected value $E(X)$ and the variance $Var(X)$. ← Statistics 1 Sections 6.4, 6.5

If two distributions X_1 and X_2 are independent, then:

$$E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$$

and

$$Var(aX_1 + bX_2) = a^2Var(X_1) + b^2Var(X_2)$$

You are not required to derive these in an exam.

We can apply these ideas to n independent distributions, $X_1, X_2, \ldots X_n$ to give:

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

and

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

We carry out a single trial X_1 from a binomial distribution, with probability of success = p

x	0	1
$P(X_1 = x)$	$1 - p$	p

This is called a **Bernoulli trial**. Using the techniques from Statistics 1, we can find $E(X_1)$ and $Var(X_1)$.

$$E(X_1) = 0 \times (1 - p) + 1 \times p = p$$

$$\begin{aligned} Var(X_1) &= E(X_1^2) - \mu^2 \\ &= 0^2 \times (1 - p) + 1^2 \times p - \mu^2 \\ &= p - p^2 = p\,(1 - p) \end{aligned}$$

A binomial distribution $X \sim B(n, p)$ is a sum of n independent Bernoulli trials, $X = X_1 + X_2 + \ldots + X_n$ and so

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

$$\begin{aligned} E(X) &= p + p + \ldots + p \\ &= np \end{aligned}$$

and

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

$$\begin{aligned} Var(X) &= p(1 - p) + p(1 - p) + \ldots + p(1 - p) \\ &= np(1 - p) \end{aligned}$$

- **If X is a binomial random variable with $X \sim B(n, p)$, then:**
  - **Mean of $X = E(X) = \mu = np$**
  - **Variance of $X = Var(X) = \sigma^2 = np(1 - p)$**

Example 7 SKILLS INTERPRETATION

A fair, five-sided spinner with segments numbered 1, 2, 3, 4 and 5 is spun 20 times. The random variable X represents the number of 5s obtained.

a Find the mean and variance of X.

b Find $P(X < \mu - \sigma)$.

X represents the number of 5s obtained in 20 spins of the spinner

Define the random variable carefully.

$X \sim B(20, 0.2)$

The value of p is $\frac{1}{5} = 0.2$, as it is a five-sided spinner and assumed to be fair.

a $E(X) = \mu = np = 20 \times 0.2 = 4$

$Var(X) = \sigma^2 = np(1 - p)$
$= 20 \times 0.2 \times 0.8 = 3.2$

b $\sigma = \sqrt{3.2} = 1.789$ (3 d.p.)

$P(X < \mu - \sigma) = P(X < (4 - 1.789))$

$= P(X < 2.211)$

$= P(X \leqslant 2)$

$= 0.2061$

As X can only take integer values, $P(X < 2.211) = P(X \leqslant 2)$.

$P(X \leqslant 2)$ can be found using binomial tables. Remember that $p = 0.2$, $n = 20$.

Example 8

SKILLS INTERPRETATION

A company produces a certain type of delicate component. The probability of any one component being faulty is p. The probability of obtaining at least one faulty component in a sample of four components is 0.3439.

The company produces 600 components in a day.

Find the mean and variance of the number of faulty components produced per day.

X represents the number of faulty components in a sample of four components

$X \sim B(4, p)$

$P(X \geqslant 1) = 0.3439$

$1 - P(X = 0) = 0.3439$

$P(X = 0) = 0.6561$

$(1 - p)^4 = 0.6561$

$1 - p = 0.9$

$p = 0.1$

Y represents the number of components produced in a day

$Y \sim B(600, 0.1)$

Mean $= E(Y) = np = 600 \times 0.1 = 60$

Variance $= \text{Var}(Y) = np(1 - p)$

$= 600 \times 0.1 \times 0.9 = 54$

Define your random variable.

To answer this question we need first to be able to find a value for p.

The probability of obtaining at least one faulty item in a sample of four components is 0.3439

$P(X = 0) = \binom{4}{0} p^0 (1 - p)^4 = 1 \times 1 \times (1 - p)^4 = (1 - p)^4$

Define a new random variable using the value of p obtained earlier.

Exercise 1C

SKILLS PROBLEM-SOLVING

1 X is the random variable such that $X \sim B(12, 0.7)$. Find:

a $E(X)$ **b** $\text{Var}(X)$

2 X is the random variable such that $X \sim B(n, 0.4)$ and $E(X) = 3.2$. Find:

a the value of n **b** $P(X = 5)$ **c** $P(X \leqslant 2)$

3 X is the random variable such that $X \sim B(10, p)$ and $\text{Var}(X) = 2.4$. Find the two possible values of p.

4 X is the random variable such that $X \sim B(15, p)$ and $\text{Var}(X) = 2.4$. Find the two possible values of p.

5 X is the random variable such that $X \sim B(n, p)$, $E(X) = 4.8$ and $Var(X) = 2.88$.
Find the values of n and p.

E/P 6 The probability of obtaining a head when a biased coin is tossed is p, where $p < \frac{1}{2}$.
An experiment consists of tossing the coin 20 times and recording the number of heads.
In a large number of experiments the variance of the number of heads is found to be 4.2.

a Estimate the value of p. **(2 marks)**

b Hence estimate the probability that exactly 7 heads are recorded during a particular experiment. **(2 marks)**

P 7 The probability that a salesman gets a reply when he knocks on the door of a house is 0.65.

a Find the probability that in a street of 10 houses he receives:

i exactly 5 replies

ii at least 5 replies.

b **i** How many houses should he visit such that the random variable X = 'number of replies' has a mean of 78?

ii What is the variance in this case?

8 A company produces and sells boxes of chocolate biscuits. When a box is opened, the probability of a biscuit being broken is 0.04

a Find the probability that in a box of 48 biscuits, at least two are broken.

The company produces 120 boxes of biscuits per day.

b Find the mean and variance of the number of boxes which contain at least two broken biscuits.

E 9 The random variable X is such that $X \sim B(5, p)$. Given that $P(X \geqslant 1) = 0.83193$, find:

a the value of p **(3 marks)**

b $E(X)$ and $Var(X)$. **(2 marks)**

P 10 A biased dice is thrown 5 times and the number of sixes is noted. The experiment is conducted 500 times. The results are shown in the table.

Number of sixes	0	1	2	3	4	5
Frequency	163	208	98	28	3	0

A student wishes to show if the data can be modelled by a binomial distribution.

a Calculate the mean and variance of the number of sixes in 5 throws of the dice.

b Based on the mean of the data, estimate the probability p of getting a six with this dice.

c Using the value found in part **b**, calculate the expected frequencies of 0, 1, 2, 3, 4 and 5 sixes in 500 experiments, using a binomial distribution with parameters $n = 5$ and p. Comment on the student's suggestion.

d How does the variance of the data support the use of a binomial distribution?

Challenge

SKILLS CREATIVITY

By writing out the probability distribution table for $X \sim B(3, p)$, show that:

a $E(X) = 3p$

b $Var(X) = 3p(1 - p)$

Chapter review 1

(E) **1** The discrete random variable $X \sim B(30, 0.75)$. Find:

a $P(X = 20)$ **(1 mark)**

b $P(X \leqslant 13)$ **(1 mark)**

c $P(11 < X \leqslant 25)$ **(2 marks)**

(P) **2** A coin is biased so that the probability of a head is $\frac{2}{3}$. The coin is tossed repeatedly. Find the probability that:

a the first tail will occur on the sixth toss

b in the first 8 tosses there will be exactly 2 tails.

(P) **3** Records kept at a hospital show that 3 out of every 10 patients who visit the Accident and Emergency department have to wait more than half an hour to be seen by a doctor. Find, to 3 decimal places, the probability that of the first 12 patients who come to the Accident and Emergency department:

a none of them wait more than half an hour

b more than 2 patients wait more than half an hour.

(E/P) **4** **a** State clearly the conditions under which it is appropriate to assume that a random variable has a binomial distribution. **(2 marks)**

A door-to-door salesman tries to persuade people to have a certain type of window installed. The probability that he is successful in selling the window is 0.05.

b Find the probability that he will have at least 2 successes out of the first 10 houses he visits. **(2 marks)**

c Calculate the fewest number of houses he must visit so that the probability of his getting at least one success is more than 0.99. **(4 marks)**

(P) **5** A completely unprepared student is given a test with 10 true-or-false questions. Assuming that the student answers all the questions at random:

a find the probability that the student gets all the answers correct.

It is decided that a pass will be awarded for 8 or more correct answers.

b Find the probability that the student passes the test.

(P) **6** A six-sided dice is biased. When the dice is thrown, the number 5 is twice as likely to appear as any other number. All other numbers are equally likely to appear. The dice is thrown repeatedly. Find the probability that:

a the first 5 will occur on the sixth throw

b in the first eight throws there will be exactly three 5s.

(E/P) **7** A manufacturer produces large quantities of chairs. It is known from previous records that 15% of these chairs are green. A random sample of 10 chairs is taken.

a Define a suitable distribution to model the number of green chairs in this sample. **(1 mark)**

b Find the probability of at least 5 green chairs in this sample. **(3 marks)**

c Find the probability of exactly 2 green chairs in this sample. **(3 marks)**

(E/P) **8** A bag contains a large number of beads of which 45% are yellow. A random sample of 20 beads is taken from the bag. Use the binomial distribution to find the probability that the sample contains:

a fewer than 12 yellow beads **(2 marks)**

b exactly 12 yellow beads. **(3 marks)**

(E/P) **9** An archer hits the bullseye with probability 0.6. She shoots 20 arrows at the target.

a Find the probability that she hits the bullseye with at least 50% of her arrows. **(3 marks)**

She shoots 12 sets of 20 arrows.

b Find the probability that she hits the bullseye with at least 50% of her arrows in 7 of the 12 sets of 20 arrows. **(2 marks)**

c Find the probability that she hits the bullseye with at least 50% of her arrows in fewer than 6 sets of 20 arrows. **(2 marks)**

(E) **10** The probability of damaged seeds successfully growing is 0.075. Find the probability that, in a batch of 20 randomly selected damaged seeds, the number that will grow is:

a **i** exactly 2 **(2 marks)**

ii more than 4. **(2 marks)**

A second random sample of 80 damaged seeds is selected.

b Find the mean and variance for this distribution. **(3 marks)**

(E) **11** A receptionist transfers incoming telephone calls to rooms within a hotel. The probability of the caller being connected to the wrong room is 0.02.

a Find the probability that more than 1 call in 10 **consecutive** calls is connected to the wrong room. **(3 marks)**

The receptionist receives 500 calls each day for guests in the hotel.

b Find the mean and variance of the number of wrongly connected calls. **(2 marks)**

(E) **12** A disease occurs in 2.5% of a population.

a Find the probability of exactly 2 people having the disease in a random sample of 10 people. **(2 marks)**

b Find the mean and variance of the number of people with the disease in a random sample of 120 people. **(2 marks)**

13 Let $X \sim B(2n, 0.25)$

Find the smallest value of n such that $P(X \geqslant n) \leqslant 0.1$

Challenge

SKILLS INNOVATION

A driving theory test has 50 questions. Each question has four answers, only one of which is correct.

Li Wei is certain she got 32 answers correct, but she guessed the remaining answers. She needs to get 43 correct answers to pass the test.

Find the probability that Li Wei passed the test.

Summary of key points

1 You can model X with a binomial distribution, $B(n, p)$, if:

- there are a fixed number of trials, n
- there are two possible outcomes (success or failure)
- there is a fixed probabillity of success, p
- the trials are independent of each other.

2 If a random variable X has the binomial distribution $B(n, p)$ then its probability is given by:

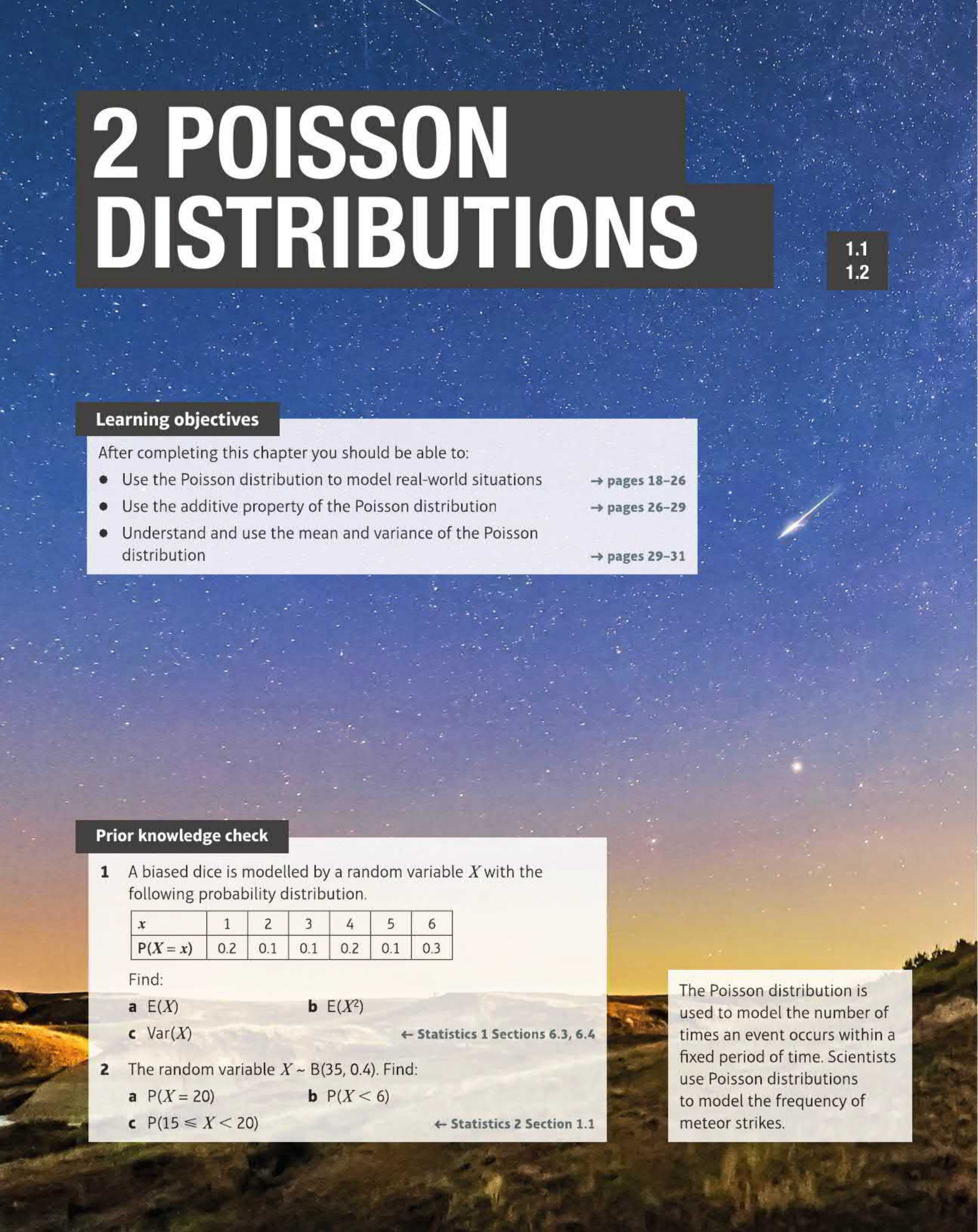
$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}$$

3 If X has a binomial distribution with $X \sim B(n, p)$, then:

- Mean of $X = E(X) = \mu = np$
- Variance of $X = \text{Var}(X) = \sigma^2 = np(1 - p)$

2 POISSON DISTRIBUTIONS

1.1
1.2



Learning objectives

After completing this chapter you should be able to:

- Use the Poisson distribution to model real-world situations → pages 18–26
- Use the additive property of the Poisson distribution → pages 26–29
- Understand and use the mean and variance of the Poisson distribution → pages 29–31

Prior knowledge check

1 A biased dice is modelled by a random variable X with the following probability distribution.

x	1	2	3	4	5	6
$P(X = x)$	0.2	0.1	0.1	0.2	0.1	0.3

Find:

a $E(X)$ **b** $E(X^2)$

c $Var(X)$ ← **Statistics 1 Sections 6.3, 6.4**

2 The random variable $X \sim B(35, 0.4)$. Find:

a $P(X = 20)$ **b** $P(X < 6)$

c $P(15 \leqslant X < 20)$ ← **Statistics 2 Section 1.1**

The Poisson distribution is used to model the number of times an event occurs within a fixed period of time. Scientists use Poisson distributions to model the frequency of meteor strikes.

2.1 The Poisson distribution

In the binomial expansion, there are a fixed number of trials with only two possible outcomes for each trial – success and failure. We count the number of successes.

Let us consider a different situation.

A small company is looking at how many complaints it receives by telephone each day. The company has just one phone line. Data is collected over 100 days, as shown in this table:

Number of calls per day	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	4	17	25	23	18	10	2	0	0	0	0	0	1

The phone calls:

- occur **singly** in time – only one phone call occurs at any given time
- are random
- occur independently in a given timeframe.

There are 6 or fewer calls per day, except one day when there were 12. In theory, the maximum number of calls could be infinite, but numbers much bigger than the mean are not anticipated.

We analyse these data further by calculating the mean and variance.

For these data, $\sum fx = 282$, $\sum fx^2 = 1078$ and $n = 100$

The mean number of calls per day is therefore $\frac{\sum fx}{n} = \frac{282}{100} = 2.82$

And the variance is $\frac{\sum fx^2}{n} - \bar{x}^2 = \frac{1078}{100} - 2.82^2 = 2.8276$

Here, the mean and variance are very close in value.

In this situation, the data can be modelled by a Poisson distribution.

For a Poisson distribution, the mean and variance are assumed to be the same and so there is only one parameter, the rate at which the calls are made per day. This is equivalent to the mean and is represented by λ.

- Letting $X \sim \text{Po}(\lambda)$, the Poisson distribution is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ where } x = 0, 1, 2, \ldots$$

We can confirm this by looking at the data above:

Let $X \sim \text{Po}(2.76)$

Number of calls per day	0	1	2	3	4	5	6	7	8	9	10	11	12
Relative frequency	0.04	0.17	0.25	0.23	0.18	0.1	0.02	0	0	0	0	0	0.01
$P(X = x)$	0.063	0.175	0.241	0.222	0.153	0.084	0.039	0.015	0.005	0.002	0.000	0.000	0.001

To show that this is a probability distribution, $\sum P(X = x) = 1$, consider each term:

$$\sum P(X = x) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + \dots$$
$$= e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!} + \dots$$

This can be factorised to:

$$\sum P(X = x) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots\right)$$

The term in the brackets is the series expansion of e^{λ}:

$$e^{\lambda} = \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots\right)$$

and so:

$$\sum P(X = x) = e^{-\lambda}e^{\lambda}$$
$$= 1 \text{ as required}$$

Links You are not required to derive these in an exam. ← Further Pure 2 Section 7.2

Example 1 SKILLS INTERPRETATION

The random variable $X \sim \text{Po}(2.1)$. Find:

a $P(X = 3)$

b $P(X \geqslant 1)$

c $P(1 < X \leqslant 4)$

a $P(X = 3) = \dfrac{e^{-2.1} \times 2.1^3}{3!}$

$= 0.189011\dots$

$= 0.1890$ (4 d.p.)

Use the formula $P(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ with $x = 3$ and $\lambda = 2.1$

You can work this out using the Poisson probability distribution function on your calculator.

b $P(X \geqslant 1) = 1 - P(X = 0)$

$= 1 - e^{-2.1}$

$= 1 - 0.1224\dots$

$= 0.8775$ (4 d.p.)

X can only take positive integer values.

Round probabilities to 4 decimal places.

c $P(1 < X \leqslant 4)$

$= P(X = 2) + P(X = 3) + P(X = 4)$

$= \dfrac{e^{-2.1} \times 2.1^2}{2!} + \dfrac{e^{-2.1} \times 2.1^3}{3!} + \dfrac{e^{-2.1} \times 2.1^4}{4!}$

$= 0.2700\dots + 0.1890\dots + 0.0992\dots$

$= 0.5583$ (4 d.p.)

Add together all the possible probabilities. The positive integers that satisfy the inequality are 2, 3 and 4.

Online Explore the Poisson distribution using GeoGebra.

Exercise 2A

1 The discrete random variable $X \sim \text{Po}(2.5)$. Find:
a $P(X = 3)$ **b** $P(X > 1)$ **c** $P(1 < X \leqslant 3)$

2 The discrete random variable $X \sim \text{Po}(3.1)$. Find:
a $P(X = 4)$ **b** $P(X \geqslant 2)$ **c** $P(1 \leqslant X \leqslant 4)$

3 The discrete random variable $X \sim \text{Po}(4.2)$. Find:
a $P(X = 2)$ **b** $P(X \leqslant 3)$ **c** $P(3 \leqslant X \leqslant 5)$

4 The discrete random variable $X \sim \text{Po}(0.84)$. Find:
a $P(X = 1)$ **b** $P(X \geqslant 1)$ **c** $P(1 < X \leqslant 3)$

(P) 5 The discrete random variable $X \sim \text{Po}(\lambda)$. Given that $P(X = 2) = P(X = 3)$, find λ.

(P) 6 The discrete random variable $X \sim \text{Po}(\lambda)$. Given that $P(X = 4) = 3 \times P(X = 2)$, find λ.

Calculations involving the Poisson distribution can often be made simpler by using the cumulative distribution tables given on page 144. These tables will be given in the *Mathematical Formulae and Statistical Tables* booklet in your exam. These will give you $P(X \leqslant x)$ for values of λ between 0 and 10, in steps of 0.5, and for values of x from 0 to 10.

Example 2 SKILLS INTERPRETATION

The random variable $X \sim \text{Po}(5)$. Find, using tables:
a $P(X \leqslant 3)$ **b** $P(X \geqslant 2)$ **c** $P(1 \leqslant X \leqslant 4)$

a $P(X \leqslant 3) = 0.2650$

b $P(X \geqslant 2) = 1 - P(X \leqslant 1) = 1 - 0.0404$
$= 0.9596$

c $P(1 \leqslant X \leqslant 4) = P(X \leqslant 4) - P(X \leqslant 0)$
$= 0.4405 - 0.0067$
$= 0.4338$

$\lambda =$	0.5	4.5	5.0
$x = 0$	0.6065	0.0111	0.0067
1	0.9098	0.0611	0.0404
2	0.9856	0.1736	0.1247
3	0.9982	0.3423	0.2650
4	0.9998	0.5321	0.4405

Be careful. The inequality is $X \geqslant 2$ so you need to work out $1 - P(X \leqslant 1)$.

Example 3 SKILLS PROBLEM-SOLVING

The random variable $X \sim \text{Po}(7.5)$. Find the values of a, b and c such that:
a $P(X \leqslant a) = 0.2414$ **b** $P(X < b) = 0.5246$ **c** $P(X \geqslant c) = 0.3380$

a $P(X \leqslant a) = 0.2414$
so $a = 5$

Use tables with $\lambda = 7.5$
$P(X \leqslant 5) = 0.2414$

b $P(X < b) = P(X \leqslant b - 1) = 0.5246$
so $b - 1 = 7$
$b = 8$

c $P(X \geqslant c) = 1 - P(X \leqslant c - 1) = 0.3380$
so $P(X \leqslant c - 1) = 1 - 0.3380$
$= 0.6620$
so $c - 1 = 8$
$c = 9$

Use tables with $\lambda = 7.5$
$P(X \leqslant 7) = 0.5246$

Use tables with $\lambda = 7.5$
$P(X \leqslant 8) = 0.6620$

Exercise 2B SKILLS PROBLEM-SOLVING

Use the Poisson cumulative distribution tables to answer these questions.

1 The discrete random variable $X \sim \text{Po}(5.5)$. Find:
a $P(X \leqslant 3)$ **b** $P(X \geqslant 6)$ **c** $P(3 \leqslant X \leqslant 7)$

2 The discrete random variable $X \sim \text{Po}(10)$. Find:
a $P(X \geqslant 8)$ **b** $P(7 \leqslant X \leqslant 12)$ **c** $P(4 < X < 9)$

3 The discrete random variable $X \sim \text{Po}(3.5)$. Find:
a $P(X \geqslant 2)$ **b** $P(3 \leqslant X \leqslant 6)$ **c** $P(2 < X \leqslant 5)$

4 The discrete random variable $X \sim \text{Po}(4.5)$. Find:
a $P(X \geqslant 5)$ **b** $P(3 < X \leqslant 5)$ **c** $P(1 \leqslant X < 7)$

5 The discrete random variable $X \sim \text{Po}(8)$. Find the values of a, b, c and d such that:
a $P(X \leqslant a) = 0.3134$ **b** $P(X \leqslant b) = 0.7166$ **c** $P(X < c) = 0.0996$ **d** $P(X > d) = 0.8088$

6 The discrete random variable $X \sim \text{Po}(3.5)$. Find the values of a, b, c and d such that:
a $P(X \leqslant a) = 0.8576$ **b** $P(X > b) = 0.6792$ **c** $P(X \leqslant c) \geqslant 0.95$ **d** $P(X > d) \leqslant 0.005$

2.2 Modelling with the Poisson distribution

You need to be able to recognise situations that can be modelled with a Poisson distribution. The Poisson distribution is used to model the number of times, X, that a particular event occurs within a given **interval** of time or space.

- In order for the Poisson distribution to be a good model, the events must occur:
  - independently
  - singly, in space or time
  - at a constant average rate so that the mean number in an interval is **proportional** to the length of the interval.

Notation We can see that in the example given at the start of the chapter, the phone calls occur singly in time, are random, and occur independently in a given time frame. In fact, these are the requirements to model using a Poisson distribution.

The parameter λ in the Poisson distribution is the average number of times that the event will occur in a single interval.

Examples of where a Poisson distribution might be appropriate are:

- the number of radioactive particles being emitted (sent out) by a certain source in a 5-minute period
- the number of telephone calls to a switchboard in a 10-minute interval
- the number of spelling mistakes on a page of a newspaper
- the number of cars passing the front of a school in a 3-minute interval
- the number of raisins in a fruitcake.

Example 4

An internet service provider has a large number of users regularly connecting to the internet. On average, 4 users every hour fail to connect to the internet on their first attempt.

a Give two reasons why a Poisson distribution might be a suitable model for the number of failed connections every hour.

b Find the probability that in a randomly chosen hour:

i 2 users fail to connect on their first attempt

ii more than 6 users fail to connect on their first attempt.

c Find the probability that in a randomly chosen 90-minute period:

i 5 users fail to connect on their first attempt

ii fewer than 7 users fail to connect on their first attempt.

a Failed connections occur singly and at a constant rate of 4 users per hour.

b X represents the number of failed connections in one hour

$X \sim \text{Po}(4)$

i $P(X = 2) = 0.1465$

ii $P(X > 6) = 1 - P(X \leqslant 6)$

$= 1 - 0.88932...$

$= 0.1107$ (4 d.p.)

Define your random variable, and write down the model you are using.

Use the tables, or your calculator, with $\lambda = 4$ to find $P(X \leqslant 6)$.

c Y represents the number of failed connections in 90 minutes

$Y \sim \text{Po}(6)$

i $P(Y = 5) = 0.1606$

ii $P(Y < 7) = P(Y \leqslant 6)$

$= 0.6063$ (4 d.p.)

Problem-solving

Because the failures occur at a **constant average rate**, the value of the parameter λ will be $\frac{90}{60} \times 4 = 6$ for a 90-minute period.

Example 5 SKILLS PROBLEM-SOLVING; INTERPRETATION

The number of patients visiting a clinic that treats insect bites can be modelled as a Poisson distribution with a rate of 3 patients per day.

a Find the probability that there are more than 4 patients on a given day.

b An extra doctor is required at the clinic if, within a 5 day period, at least 4 of the days have more than 4 patients. Find the probability that an extra doctor is required.

Let X represent the number of patients visiting the clinic per day: $X \sim \text{Po}(3)$ — Always define the variable.

a We are finding:

$$P(X > 4) = 1 - P(X \leqslant 4)$$
$$= 1 - 0.8153$$ — Using tables
$$= 0.1847$$

b We are now talking about the number of days in which there are more than 4 patients visiting the clinic.

Let Y represent the number of days when there are more than 4 patients visiting the clinic in a 5 day period. — Define the variable. This can now be modelled as a binomial distribution.

$Y \sim \text{B}(5, 0.1847)$

We are finding:

$$P(Y \geqslant 4) = P(Y = 4) + P(Y = 5)$$
$$= \binom{5}{4}(0.1847)^4(0.8153) + \binom{5}{5}(0.1847)^5$$ — Using the formula
$$= 0.004744113 + 0.000214949$$
$$= 0.00496$$ — Write the answer to 3 s.f.

Exercise 2C SKILLS PROBLEM-SOLVING

1 The maintenance department at a school receives requests for replacement light bulbs at a rate of 3 per week. The number of requests, X, in a given week is modelled as $X \sim \text{Po}(3)$.

a Find the probability that, in a randomly chosen week, the number of requests for replacement light bulbs is:

i exactly 4 **ii** more than 5.

b Find the probability that, in a randomly chosen two week period, the number of requests for replacement light bulbs is:

i exactly 6

ii no more than 4.

Hint The number of requests, Y, in a given two week period can be modelled as $Y \sim \text{Po}(6)$.

2 A botanist suggests that the number of weeds growing in a field can be modelled by a Poisson distribution.

a Write down two conditions that must apply for this model to be suitable.

Assuming this model and that weeds occur at a rate of 1.3 per m^2, find:

b the probability that, in a randomly chosen plot of $4\,m^2$, there will be fewer than 3 weeds

c the probability that, in a randomly chosen plot of $5\,m^2$, there will be more than 8 weeds.

Problem-solving

The number of weeds, X, in a plot of $4\,m^2$ can be modelled as $X \sim Po(4 \times 1.3)$, i.e. $X \sim Po(5.2)$.

3 An electronics company manufactures a component for use in computer hardware. At the end of the manufacturing process, each component is checked to see if it is faulty. Faulty components are detected (discovered) at a rate of 2.5 per hour.

a Suggest a suitable model for the number of faulty components detected per hour.

b Describe, in the context of this question, two assumptions you have made in part **a** for this model to be suitable.

c Find the probability of 2 faulty components being detected in a 1-hour period.

d Find the probability of at least 6 faulty components being detected in a 3-hour period.

e Find the probability of at least 7 faulty components being detected in a 4-hour period.

4 A call-centre agent handles telephone calls at a rate of 15 per hour.

a Find the probability that, in a randomly selected 20-minute interval, the agent handles:
i exactly 4 calls **ii** more than 8 calls.

b Find the probability that, in a randomly selected 30-minute interval, the agent handles:
i at least 6 calls **ii** no more than 10 calls.

5 The average number of cars crossing over a bridge is 180 per hour.
Assuming a Poisson distribution, find the probability that:

a more than 5 cars will cross in any given minute

b fewer than 4 cars will cross in any 2-minute period.

6 A café serves breakfast every morning. Customers arrive for breakfast at random at an average rate of one every 4 minutes.
Find the probability that on a Friday morning between 10 am and 10:20 am:

a fewer than 3 customers arrive for breakfast

b more than 10 customers arrive for breakfast.

E/P **7** An estate agent has been selling houses at a rate of 1.8 per week.

SKILLS ADAPTIVE LEARNING

a Find the probability that in a particular week she sells:

i no houses **ii** 3 houses **iii** at least 3 houses. **(6 marks)**

The estate agent meets her weekly target if she sells at least 3 houses in one week.

b Find the probability that over a period of 4 consecutive weeks she meets her weekly target exactly once. **(3 marks)**

Problem-solving

Use a binomial model for part **b**.

← **Statistics 2 Section 1.1**

E **8** Patients arrive at a hospital Accident and Emergency department at random at a rate of 5 per hour.

a Find the probability that, during any 30-minute period, the number of patients arriving at the Accident and Emergency department is:

i exactly 4 **ii** at least 3. **(5 marks)**

A patient arrives at 11:00 am.

b Find the probability that the next patient arrives before 11:15 am. **(3 marks)**

E **9** The elevator in an apartment building breaks down at random at a mean rate of three times per four-week period.

a Find the probability that the elevator breaks down:

i at least once in one week

ii exactly twice in one week. **(5 marks)**

In one particular week, the elevator broke down twice.

b Write down the probability that the elevator will break down at some point in the next week. Give a reason for your answer. **(2 marks)**

E/P **10** Flaws (mistakes) occur at random in a particular type of material at a mean rate of 1.5 per 50 m.

a Find the probability that, in a randomly chosen 50 m length of this material, there will be exactly 3 flaws. **(2 marks)**

This material is sold in rolls of length 200 m.

b Find the probability that a single roll has fewer than 4 flaws. **(3 marks)**

Priya buys 5 rolls of this material.

c Find the probability that at least two of these rolls will have fewer than 4 flaws. **(5 marks)**

E/P **11** A company produces chocolate chip biscuits. The number of chocolate chips per biscuit has a Poisson distribution with mean 5.

a Find the probability that one of these biscuits, selected at random, contains fewer than 3 chocolate chips. **(2 marks)**

A packet contains 6 of these biscuits, selected at random.

b Find the probability that exactly half of the biscuits in the packet contain fewer than 3 chocolate chips. **(4 marks)**

(E/P) **12** A company has buses that can only be hired for a week at a time. All hiring starts on a Sunday. During the summer, the mean number of requests for buses each Sunday is 5.

a Calculate the probability that fewer than 4 requests for buses are made on a particular Sunday in the summer. **(2 marks)**

The company wants to be at least 99% sure they can fulfil all requests on any particular Sunday.

b Calculate the number of buses the company must have to satisfy this condition. **(3 marks)**

(E/P) **13** On a typical summer day, a boat company hires out rowing boats at a rate of 9 per hour.

a Find the probability of hiring out at least 6 boats in a randomly selected 30-minute period. **(2 marks)**

The company has 8 boats and decides to hire them out for 20-minute periods.

b Show that the probability of running out of boats is less than 1%. **(3 marks)**

c Find the number of boats that the company should have in order to be 99% sure of meeting all demands if the hiring period is extended to 30 minutes. **(3 marks)**

(E/P) **14** Breakdowns on a particular machine occur at a rate of 1.5 per week.

SKILLS ADAPTIVE LEARNING

a Find the probability that no more than 2 breakdowns occur in a randomly chosen week. **(2 marks)**

b Find the probability of at least 5 breakdowns in a randomly chosen two-week period. **(3 marks)**

A maintenance firm agrees to repair breakdowns over a six-week period.
The firm will give a full refund if there are more than n breakdowns in a six-week period.
The firm wants the probability of having to pay a refund to be 5% or less.

c Find the smallest possible value of n. **(3 marks)**

2.3 Adding Poisson distributions

If two Poisson variables X and Y are independent, then the variable $Z = X + Y$ also has a Poisson distribution.

- If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$, then $X + Y \sim \text{Po}(\lambda + \mu)$

Watch out For $X + Y$ to be meaningful in this context, the random variables X and Y must both model events occurring within the same interval of time or space.

Example 6

If $X \sim \text{Po}(3.6)$ and $Y \sim \text{Po}(4.4)$, find:

a $\text{P}(X + Y = 7)$ **b** $\text{P}(X + Y \leqslant 5)$

$X + Y \sim \text{Po}(3.6 + 4.4)$ — Add the parameters.

$X + Y \sim \text{Po}(8)$

a $\text{P}(X + Y = 7) = \dfrac{e^{-8} \times 8^7}{7!} = 0.1396$ (4 d.p.)

b $\text{P}(X + Y \leqslant 5) = 0.1912$ (4 d.p.)

Use the tables, or your calculator, with $\lambda = 8$.

Example 7 SKILLS PROBLEM-SOLVING

The number of cars passing an observation point in a 5-minute interval is modelled by a Poisson distribution with mean 2. The number of other vehicles passing the observation point in a 15-minute interval is modelled by a Poisson distribution with mean 3.

Find the probability that:

a exactly 5 vehicles, of any type, pass the observation point in a 10-minute interval

b more than 8 vehicles, of any type, pass the observation point in a 15-minute interval.

a X_1 represents number of cars passing in a 10-minute interval
Y_1 represents number of other vehicles passing in a 10-minute interval
$X_1 \sim \text{Po}(4)$, $Y_1 \sim \text{Po}(2)$
$X_1 + Y_1 \sim \text{Po}(4 + 2)$
$X_1 + Y_1 \sim \text{Po}(6)$
$\text{P}(X_1 + Y_1 = 5) = 0.1606$ (4 d.p.)

Problem-solving

You need to model the number of cars passing in a 10-minute interval, and the number of other vehicles passing in a 10-minute interval. The time intervals must be the same before you can add the parameters.

b X_2 represents number of cars passing in a 15-minute interval
Y_2 represents number of other vehicles passing in a 15-minute interval
$X_2 \sim \text{Po}(6)$, $Y_2 \sim \text{Po}(3)$
$X_2 + Y_2 \sim \text{Po}(6 + 3)$
$X_2 + Y_2 \sim \text{Po}(9)$
$\text{P}(X_2 + Y_2 > 8) = 1 - \text{P}(X_2 + Y_2 \leqslant 8)$
$= 1 - 0.45565\ldots$
$= 0.5443$ (4 d.p.)

Define new random variables for the number of cars, and other types of vehicle, passing in a 15-minute interval.

This can be calculated using the tables, with $\lambda = 9$.

Exercise 2D

1 X and Y are independent random variables such that $X \sim \text{Po}(3.3)$ and $Y \sim \text{Po}(2.7)$. Find:
a $\text{P}(X + Y = 5)$ **b** $\text{P}(X + Y \leqslant 7)$ **c** $\text{P}(X + Y > 4)$

2 A and B are independent random variables such that $A \sim \text{Po}(3.25)$ and $B \sim \text{Po}(4.25)$. Find:
a $\text{P}(A + B = 7)$ **b** $\text{P}(A + B \leqslant 5)$ **c** $\text{P}(A + B > 9)$

(P) 3 X and Y are independent random variables such that $X \sim \text{Po}(2.5)$ and $Y \sim \text{Po}(3.5)$. Find:
a $\text{P}(X = 2 \text{ and } Y = 2)$
b P(both X and Y are greater than 2)
c $\text{P}(X + Y = 5)$
d $\text{P}(X + Y \leqslant 4)$

(P) **4** The number of emissions (gas that is sent out into the air) per minute from two different sources of radioactivity are modelled as independent Poisson random variables X and Y, with parameters of 3 and 5 respectively (in the same order as mentioned before).
Calculate the probability that, in a given one-minute period:

a the number of emissions from each source is at least 3

b the total number of emissions from the two sources is no more than 6.

(P) **5** During a weekday at a certain point on a road, cars pass by at a rate of 24 per minute, while vans pass by at a rate of 8 per minute.

a Find the probability that, in any 15-second period:

i at least 4 of each type of vehicle passes by

ii the total number of cars and vans that pass by is no more than 9.

b Write down one assumption that you have made in your calculations.

(E) **6** A taxi company supplies two particular organisations independently.
Company A orders taxis at a rate of 1.25 cars per day.
Company B orders taxis at a rate of 0.75 cars per day.

a On a given day, find the probability that 2 cars are ordered by company A. **(2 marks)**

b On a given day, find the probability that the total number of cars ordered by both companies is 2. **(2 marks)**

c In a given 5-day working week, find the probability that the total number of cars ordered by both companies is less than 10. **(2 marks)**

(E/P) **7** A restaurant has two coffee machines, C and D. Machine C breaks down at a rate of 0.1 times per week while, independently, machine D breaks down at a rate of 0.05 times per week.
Find the probability that, in a 12-week period:

a machine C breaks down at least once **(2 marks)**

b each machine breaks down at least once **(3 marks)**

c there will be a total of 3 breakdowns. **(2 marks)**

(E) **8** An office worker receives internal calls at a rate of one call every 5 minutes, and external calls at a rate of two calls every 5 minutes.
Calculate the probability that the total number of calls is:

a three, in a 4-minute period **(2 marks)**

b at least two, in a 2-minute period **(2 marks)**

c no more than five, in a 10-minute period. **(2 marks)**

(E/P) **9** An office is situated on three floors of a building. There is a photocopier on each floor. The ground-floor photocopier breaks down at a rate of 0.4 times per week, the first-floor photocopier breaks down at a rate of 0.2 times per week and the second-floor photocopier breaks down at a rate of 0.8 times per week. Find the probability, in a given week, that:

a each photocopier will break down exactly once **(3 marks)**

b at least one photocopier breaks down **(3 marks)**

c there will be a total of two breakdowns. **(2 marks)**

E/P **10** During the working day, the emails arriving to the account of a company director are organised into three types: personal, business and advertising. Personal emails arrive at a mean rate of 1.8 per hour, business emails arrive at a mean rate of 3.7 per hour and advertising emails arrive at a mean rate of 1.5 per hour. Find the probability that she receives:

a at least one of each type of email during a 30-minute period of the working day **(3 marks)**

b more than 50 emails in an 8-hour working day. **(3 marks)**

c Find the probability that she receives more than 50 emails on exactly two days out of a 5-day working week. **(3 marks)**

Hint Use a binomial model for part **c**.
← **Statistics 2 Section 1.1**

Challenge

$X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$. The random variable $Q = X + Y$.

a Show that $\text{P}(Q = 0) = \text{e}^{-(\lambda + \mu)}$

b Show that $\text{P}(Q = 1) = (\lambda + \mu)\text{e}^{-(\lambda + \mu)}$

2.4 Mean and variance of a Poisson distribution

It can be shown that if the random variable X has a Poisson distribution with parameter λ, then the mean and variance of X are both equal to λ.

- If $X \sim \text{Po}(\lambda)$
  - Mean of $X = \text{E}(X) = \lambda$
  - Variance of $X = \text{Var}(X) = \sigma^2 = \lambda$

Hint If you wish to prove this, which is beyond the scope of this course, you can research probability generating functions which will then lead you to a proof.

The fact that the mean is equal to the variance is an important property of a Poisson distribution. The presence or absence of this property can be a useful indicator of whether or not a Poisson distribution is a suitable model for a particular situation.

Example 8 SKILLS REASONING

A botanist counts the number of daisies, x, in each of 80 randomly selected squares within a field. The results are summarised below.

$$\Sigma x = 295, \ \Sigma x^2 = 1386$$

a Calculate the mean and the variance of the number of daisies per square for the 80 squares. Give your answers to 2 decimal places.

b Explain how the answers from part **a** support the choice of a Poisson distribution as a model.

c Using a suitable value for λ, estimate the probability that exactly 3 daisies will be found in a randomly selected square.

a Mean $= \bar{x} = \frac{\Sigma x}{80} = \frac{295}{80} = 3.69$ (2 d.p.)

Variance $= \sigma^2 = \frac{\Sigma x^2}{80} - \bar{x}^2$

$= \frac{1386}{80} - \left(\frac{295}{80}\right)^2$

$= 3.73$ (2 d.p.)

b Both the mean and the variance are 3.7 correct to 1 decimal place. The fact that the mean is close to the variance supports the choice of a Poisson distribution as a model.

c Using $\lambda = 3.7$

X = the number of daisies per square

$X \sim \text{Po}(3.7)$

$\text{P}(X = 3) = 0.2087$ (4 d.p.)

Use $\lambda = 3.7$, which is the mean and variance from part **b**.

This can be calculated using the tables, or your calculator, with $\lambda = 3.7$

Exercise 2E

SKILLS REASONING

1 A student is investigating the number of cherries in a fruitcake. A random sample of 100 fruitcakes is taken and the results are summarised as:

$\Sigma x = 143, \Sigma x^2 = 347$

a Calculate the mean and the variance of the data.

b Explain why the results in part **a** suggest that a Poisson distribution may be a suitable model for the number of cherries in a fruitcake.

c Using a suitable value for λ, estimate the probability that exactly 3 cherries will be found in a randomly selected fruitcake.

2 The number of cars passing a checkpoint during 200 periods of 5 minutes is recorded.

Number of cars	0	1	2	3	4	5	6	7	8	$\geqslant 9$
Frequency	7	21	30	41	36	29	21	11	4	0

a Calculate the mean and the variance of the data.

b Explain why the results in part **a** suggest that a Poisson distribution may be a suitable model for the number of cars passing the checkpoint in a 5-minute period.

c Using a suitable value for λ, estimate the probability that no more than 2 cars will pass the checkpoint in a given 5-minute period.

d Compare your answer to part **c** with the relative frequency of obtaining no more than 2 cars from the sample.

(E) **3** Tests for flaws are carried out on 120 pieces of cloth made in a textiles factory.
The results of the tests are shown in the table.

Number of flaws	0	1	2	3	4	5	6	7	$\geqslant 8$
Number of pieces	8	19	28	25	19	11	7	3	0

a Calculate the mean and the variance of the data. **(4 marks)**

b Explain why the results in part **a** suggest that a Poisson distribution may be a suitable model for the number of flaws on a piece of cloth. **(1 mark)**

The factory produces 10 000 pieces of cloth each week, and wants to estimate the number that will have 8 or more flaws.

c Explain why an estimate based on the observed relative frequencies would not be useful. **(1 mark)**

d Use a Poisson distribution to estimate the number of pieces of cloth with 8 or more flaws. **(3 marks)**

Challenge

If $X \sim \text{Po}(\lambda)$, then the distribution of X can be written as:

x	0	1	2	3	…	r	…
$\text{P}(X = x)$	$e^{-\lambda}$	$\frac{e^{-\lambda}\lambda^1}{1!}$	$\frac{e^{-\lambda}\lambda^2}{2!}$	$\frac{e^{-\lambda}\lambda^3}{3!}$		$\frac{e^{-\lambda}\lambda^r}{r!}$	

Using this distribution, show that $\text{E}(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Chapter review 2

(E/P) **1** On a particular road, accidents occur at a rate of 0.7 per month.
Find the probability of:

a no accidents in the next month **(2 marks)**

b exactly two accidents in the next 3-month period **(2 marks)**

c no accidents in exactly 2 of the next 6 months. **(3 marks)**

(E) **2** The random variable X is the number of mistakes per chapter in the first edition of a new textbook.

a State two conditions under which a Poisson distribution is a suitable model for X. **(2 marks)**

The number of mistakes per chapter has a Poisson distribution with mean 2.25.
Find the probability that:

b a randomly chosen chapter has no more than one mistake **(3 marks)**

c the total number of mistakes in two randomly chosen chapters is more than 6. **(3 marks)**

E/P **3** The random variable $Y \sim \text{Po}(\lambda)$.
Find the value of λ such that $P(Y = 5)$ is 1.25 times the value of $P(Y = 3)$. **(3 marks)**

E **4** A company receives emails at a mean rate of 3 emails every 5 minutes.

a Give two reasons why a Poisson distribution could be a suitable model for the number of emails received. **(2 marks)**

b Calculate the probability that, in a 10-minute period, the company receives:
i exactly 7 emails **(2 marks)**
ii at least 8 emails. **(2 marks)**

E **5** **a** State the conditions for which the Poisson distribution may be used as an **approximation** to the binomial distribution. **(2 marks)**

Left-handed people make up 8% of a population. A random sample of 50 people is taken from this population. The discrete random variable X represents the number of left-handed people in the sample.

b Calculate $P(X \leqslant 3)$. **(3 marks)**

c Using a Poisson approximation, estimate $P(X \leqslant 3)$. **(3 marks)**

d Calculate the percentage error in using the Poisson approximation. **(2 marks)**

E/P **6** The number of telephone calls per hour received by a small business is a random variable with distribution $\text{Po}(\lambda)$, where λ is an integer. Natalia records the number of calls, Y, received in an hour.
Given that $P(Y > 10) < 0.1$, find the largest possible value of λ. **(3 marks)**

Hint Use the Poisson distribution tables.

E/P **7** An angler is known to catch fish at a mean rate of 2 per hour. The number of fish caught by the angler in an hour follows a Poisson distribution.
The angler takes 5 fishing trips, each lasting 2 hours.
Find the probability that the angler catches at least 5 fish on exactly 3 of these trips. **(5 marks)**

Hint Find $P(X \geqslant 5)$ using a Poisson variable. Then use this value as the parameter p in a binomial model.

E/P **8** The number of cherries in a *Megan's* fruitcake follows a Poisson distribution with mean 2.5.
A *Megan's* fruitcake is to be selected at random. Find the probability that it contains:

a **i** exactly 4 cherries **(2 marks)**
ii at least 3 cherries. **(2 marks)**

Megan's fruitcakes are sold in packets of four.

b Calculate the probability that there are more than 12 cherries, in total, in a randomly selected packet of *Megan's* fruitcakes. **(3 marks)**

Eight packets of *Megan's* fruitcakes are selected at random.

c Find the probability that exactly two packets contain more than 12 cherries. **(3 marks)**

E/P **9** A car salesman sells cars at a mean rate of 6 per week.

a Suggest a suitable model to represent the number of cars sold in a randomly chosen week. Give two reasons to support your model. **(2 marks)**

b Find the probability that in any randomly chosen week the salesman sells exactly 5 cars. **(2 marks)**

c Find the probability that in a period of 4 consecutive weeks there are exactly 2 weeks in which the salesman sells exactly 5 cars. **(3 marks)**

E/P **10** Aisha and Biyu share a house. Aisha receives letters at a mean rate of 1.2 letters per day while Biyu receives letters at a rate of 0.8 letters per day. Assuming their letters are separate, calculate the probability that on a particular day:

a each receives at least 1 letter **(3 marks)**

b they receive a total of 3 letters between them. **(2 marks)**

Given that post is delivered to the house from Monday to Friday:

c find the probability that in one particular week they receive a total of three letters on at least 3 of the days. **(4 marks)**

E/P **11** An electronics shop sells desktop and laptop computers. The desktops are sold at a mean rate of 2.4 per day and the laptops are sold at a mean rate of 1.6 per day. Calculate the probability that on a particular day the shop sells:

a at least 2 desktops and at least 2 laptops **(3 marks)**

b a combined total of 6 computers. **(2 marks)**

c Calculate the probability that over a two-day period they sell a combined total of no more than 6 computers. **(3 marks)**

E/P **12** Accidents occur randomly at a roundabout at a rate of 15 every year.

a Find the probability that there will fewer than 5 accidents at the roundabout in a 6-month period. **(2 marks)**

b Find the probability that there will be at least 1 accident in a single month. **(2 marks)**

c Find the probability that there is at least 1 accident in exactly 4 months of a 6-month period. **(3 marks)**

E/P **13** An office photocopier breaks down randomly at a rate of 8 times per year.

a Find the probability that there will be exactly 2 breakdowns in the next month. **(2 marks)**

b Find the probability of at least 2 breakdowns in 3 of the next 4 months. **(3 marks)**

E **14** A holiday website receives visits at a rate of 240 per hour.

a State a distribution that is suitable to model the number of visits during a one-minute interval, and justify your choice of distribution. **(3 marks)**

Find the probability of:

b eight visits in a given minute **(2 marks)**

c at least 10 visits in 2 minutes. **(2 marks)**

E **15** The number of policies sold by a life insurance company employee each week over a 150-week period is recorded.

Number of policies sold	0	1	2	3	4	5	6	7	8
Number of weeks	10	23	35	33	24	14	7	3	1

a Calculate the mean and the variance of the data. **(3 marks)**

b Explain why the results in part **a** suggest that a Poisson distribution may be a suitable model for the number of policies sold in a week. **(1 mark)**

Challenge

During normal operational hours, planes land at an airport at an average rate of one plane every four minutes.

Given that exactly 10 planes landed at the airport between 2 pm and 3 pm, find the probability that:

a exactly 5 planes landed between 2 pm and 2:30 pm

b more than 7 planes landed between 2 pm and 2:30 pm

Summary of key points

1 If $X \sim \text{Po}(\lambda)$, then the Poisson distribution is given by:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ where } x = 0, 1, 2, 3, \ldots$$

2 In order for the Poisson distribution to be a good model, the events must occur:

- independently
- singly, in space or time
- at a constant average rate so that the mean number in an interval is proportional to the length of an interval.

3 If two Poisson variables X and Y are independent, the variable $Z = X + Y$ also has a Poisson distribution. If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$, then $X + Y \sim \text{Po}(\lambda + \mu)$

4 If X has a Poisson distribution with $X \sim \text{Po}(\lambda)$, then:

- Mean of $X = E(X) = \lambda$
- Variance of $X = \text{Var}(X) = \sigma^2 = \lambda$

3 APPROXIMATIONS

1.3
3.2

Learning objectives

After completing this chapter you should be able to:

- Use the Poisson distribution as an approximation to the binomial distribution → pages 36–39
- Approximate a binomial distribution using a normal distribution → pages 39–42
- Approximate a Poisson distribution using a normal distribution → pages 42–44
- Select appropriate distributions and solve real-life problems → pages 44–45

Prior knowledge check

1 $X \sim \text{B}(20, 0.4)$. Find:

a $\text{P}(X = 6)$ **b** $\text{P}(X \geqslant 8)$ **c** $\text{P}(3 \leqslant X \leqslant 10)$

← Statistics 2 Section 1.1

2 The probability that a plate made using a particular production process is faulty is given as 0.16. A sample of 20 plates is taken. Find:

a the probability that exactly two plates are faulty

b the probability that no more than three plates are faulty.

← Statistics 2 Section 1.1

Approximations are used in economic analysis to find answers to probability questions, without the need for lots of computation.

3.1 Using the Poisson distribution to approximate the binomial distribution

Evaluating binomial probabilities when n is large can be quite difficult and in such situations it is sometimes useful to use an approximation.

- If $X \sim \text{B}(n, p)$ and
  - n is large
  - p is small

  then X can be **approximated** by $\text{Po}(\lambda)$, where $\lambda = np$

There is no clear rule to describe what a 'large n' or a 'small p' is, but usually the value for np will be $\leqslant 10$. Generally, the larger the value of n and the smaller the value of p, the better the approximation will be. In this situation, $(1 - p)$ will be close to 1, so $\text{Var}(X) = np(1 - p)$ will be close to the mean of the distribution, $\text{E}(X) = np$. This satisfies the condition for a Poisson distribution model that the mean and variance are close.

In general, a question will state when you need to use an approximation.

Example 1

SKILLS ANALYSIS; REASONING

The random variable $X \sim \text{B}(200, 0.03)$

a Find $\text{P}(X = 4)$.

A Poisson variable $Y \sim \text{Po}(\lambda)$ is used to approximate X.

b Write down the value of λ and justify the use of a Poisson approximation in this context.

c Find $\text{P}(Y = 4)$ and comment on the accuracy of the approximation.

a $\text{P}(X = 4) = \binom{200}{4}(0.03)^4\,(0.97)^{196}$

$= 0.1338$ (4 d.p.)

b Under a Poisson approximation,

$Y \sim \text{Po}(200 \times 0.03)$, i.e $\lambda = 6$

There is no clear rule to describe what defines a 'large n' or a 'small p', but usually the value for np will be $\leqslant 10$.

You do not need to show that this inequality is true.

c $\text{P}(Y = 4) = \dfrac{e^{-6} \times 6^4}{4!} = 0.1339$ (4 d.p.)

The answer obtained from the Poisson approximation is close to the value obtained from the binomial distribution, so the approximation is accurate.

Compare your answers for parts **a** and **c**.

Example 2 SKILLS REASONING

The probability of a component produced by a certain machine being faulty is 0.007.
The number of faulty components in a batch of 1000 components is noted.

a Find the probability that exactly 6 components are faulty.

b Use a Poisson approximation to find the probability that more than 7 components are faulty.

c Explain why the approximation in part **b** is valid.

a X represents the number of faulty components in a batch of 1000

$X \sim B(1000, 0.007)$

$P(X = 6) = \binom{1000}{6} \times (0.007)^6 \times (0.993)^{994}$

$= 0.1494$ (4 d.p.)

Define the random variable.

b Under a Poisson approximation,

$X \sim Po(1000 \times 0.007)$ i.e. $X \sim Po(7)$

$P(X > 7) = 1 - P(X \leqslant 7) = 1 - 0.5987$

$= 0.4013$

This value can be calculated from tables or using your calculator.

c The approximation in part **b** is valid as n is large and p is small.

Exercise 3A SKILLS PROBLEM-SOLVING; INTERPRETATION

1 The random variable $X \sim B(100, 0.05)$

a Calculate:

i $P(X = 4)$ **ii** $P(X \leqslant 2)$

b Use a Poisson approximation to find estimates for the probabilities calculated in part **a**.

2 The random variable $X \sim B(150, 0.04)$

a Calculate:

i $P(X = 5)$ **ii** $P(X \leqslant 3)$

b Use a Poisson approximation to find estimates for the probabilities calculated in part **a**.

(P) **3** The random variable $Y \sim B(200, 0.98)$

a Calculate:

i $P(Y = 197)$ **ii** $P(Y \geqslant 198)$

b Use a Poisson approximation to find estimates for the probabilities calculated in part **a**.

Hint Create a variable $X \sim B(200, 0.02)$ which satisfies the conditions for a Poisson approximation. Hence $P(Y = 197)$ becomes $P(X = 3)$.

4 There are 800 pupils in a school.
Find the probability that exactly 4 of them have a birthday on 1 April:

a by using a binomial distribution

b by using a Poisson approximation.

c Comment on your answers to parts **a** and **b**.

Hint If X = 'number of pupils out of 800 having a birthday on 1 April', then $X \sim B(800, \frac{1}{365})$.

5 In a manufacturing process, the proportion of faulty items is 3%. For a batch of 100 items, use a Poisson approximation to find the probability that:

a there are fewer than 4 faulty items

b there are exactly 2 faulty items.

6 A medical practice tests a random sample of 180 of its patients for a certain condition which is present in 2% of the population. Using a Poisson approximation, find the probability that they find:

a one patient with the condition

b at least two patients with the condition.

7 A researcher has suggested that one in 120 people is likely to catch a particular virus.

a Assuming that a person catching the virus is independent of any other person catching it, find the probability that in a random sample of 20 people, exactly one of them catches the virus.

b Using a Poisson approximation, estimate the probability that in a random sample of 900 people, fewer than 7 catch the virus.

8 From company records, a manager knows the probability that a faulty article is produced by a particular production line is 0.025.
A random sample of 10 articles is selected from the production line.

a Find the probability that exactly one of them is faulty.

On another occasion, a random sample of 120 articles is taken.

b Using a Poisson approximation, find the probability that fewer than 4 of them are faulty.

9 A manufacturer produces large quantities of pots. 5% of the pots produced are broken.
A random sample of 10 pots is taken from the production line.

a Define a suitable distribution to model the number of broken pots in this sample.

b Find the probability that there are exactly 3 broken pots in the sample.

A new random sample of 140 pots is taken.

c Find the probability that there are between 6 and 9 (inclusive) broken pots in this sample, using a Poisson approximation.

10 The probability that a tomato plant grows over 2 metres high is 0.08. A random sample of 50 tomato plants is taken and each tomato plant's height is recorded.
Find, using a Poisson approximation, the probability that the number of tomato plants over 2 metres high is between 5 and 8 (inclusive).

(E) **11** Each cell of a certain insect contains 1200 genes. It is known that each gene has a probability 0.005 of being damaged. A cell is chosen at random.

a Suggest a suitable model for the distribution of the number of damaged genes in the cell. **(1 mark)**

b Find the mean and variance of the number of damaged genes in the cell. **(2 marks)**

c Using a Poisson approximation, find the probability that there are at most 4 damaged genes in the cell. **(3 marks)**

(E/P) **12** A machine that manufactures nails is known to produce 2.5% faulty nails. The nails are sold in packets of 200.

a Using a Poisson approximation, calculate the probability that a randomly selected packet contains more than 6 faulty nails. **(3 marks)**

A carpenter buys 6 packets of nails.

b Estimate the probability that more than half of these packets contain more than 6 faulty nails. **(4 marks)**

(E/P) **13** The probability of an electrical component being faulty is 0.0125. The component is supplied in boxes of 400.

a Using a Poisson approximation, estimate the probability that there are more than 3 faulty components in a box. **(3 marks)**

A retailer buys 5 boxes of components.

b Estimate the probability that there are more than 3 faulty components in 3 of the boxes. **(3 marks)**

Challenge

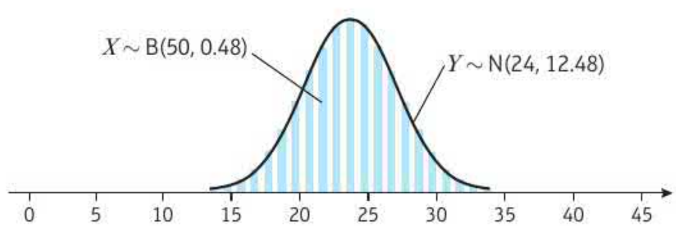
Let $X \sim B(n, p)$.

Show that if we approximate X to a Poisson distribution, then the approximation will always be an overestimate for $P(X = 0)$.

3.2 Approximating a binomial distribution

Consider the binomial random variable $X \sim B(n, p)$. It can be difficult to calculate probabilities for X when n is large. In certain circumstances you can use a normal distribution to **approximate** a binomial distribution.

Watch out The cumulative binomial tables in the formulae booklet only go up to $n = 50$.



You need to understand the conditions for which this approximation is valid, and learn the relationship between the values of n and p in $B(n, p)$ and the values of μ and σ in the normal approximation $N(\mu, \sigma^2)$.

- If n is large and p is close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where:
  - $\mu = np$
  - $\sigma = \sqrt{np(1-p)}$

Hint The approximation is valid only when p is close is to 0.5 because the normal distribution is **symmetrical**.

Example 3 SKILLS INTERPRETATION

A biased coin has P(Head) = 0.53. The coin is tossed 100 times and the number of heads X is recorded.

a Write down a binomial model for X.

b Explain why X can be approximated with a normal distribution, $Y \sim N(\mu, \sigma^2)$.

c Find the values of μ and σ in this approximation.

a $X \sim B(100, 0.53)$

b The distribution can be approximated with a normal distribution since n is large and p is close to 0.5.

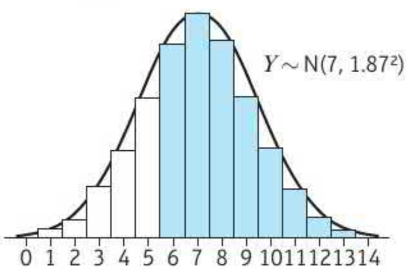
c $\mu = 100 \times 0.53 = 53$ — Use $\mu = np$

$\sigma = \sqrt{100 \times 0.53 \times (1 - 0.53)} = 4.99$ (3 s.f.) — Use $\sigma = \sqrt{np(1-p)}$

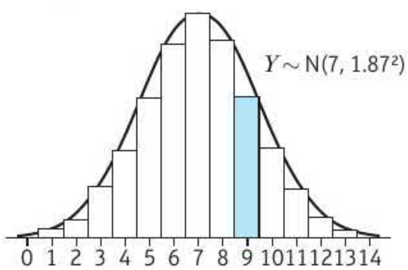
The binomial distribution is a **discrete** distribution but the normal distribution is **continuous**.

- **If you are using a normal approximation to a binomial distribution, you need to apply a continuity correction when calculating probabilities.**

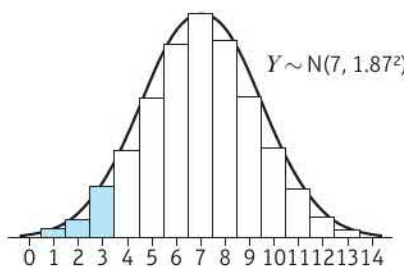
The diagrams show $X \sim B(14, 0.5)$ being approximated by $Y \sim N(7, 1.87^2)$:



$P(X > 5) \approx P(Y > 5.5)$



$P(X = 9) \approx P(8.5 < Y < 9.5)$



$P(X \leqslant 3) \approx P(Y < 3.5)$

Example 4 SKILLS PROBLEM-SOLVING

The binomial random variable $X \sim B(150, 0.48)$ is approximated by the normal random variable $Y \sim N(72, 6.12^2)$

Use this approximation to find:

a $P(X \leqslant 70)$ **b** $P(80 \leqslant X < 90)$

Watch out Remember to apply the continuity correction. You are interested in values of the **discrete** random variable X that are less than **or equal to** 70, so you need to consider values less than 70.5 for the **continuous random variable** Y.

a $P(X \leqslant 70) \approx P(Y < 70.5) = 0.4032$ (4 d.p.)

b $P(80 \leqslant X < 90) \approx P(79.5 < Y < 89.5)$
$= 0.9979 - 0.8898$
$= 0.1081$ (4 d.p.)

For values of X less than 90 consider values of Y less than 89.5

Example 5 SKILLS ADAPTIVE LEARNING

For a particular type of flower bulb, 55% will produce yellow flowers.
A random sample of 80 bulbs is planted.

Calculate the percentage error incurred when using a normal approximation to estimate the probability that there are exactly 50 yellow flowers.

Let X represent the number of bulbs producing yellow flowers in a sample of 80.
Then $X \sim B(80, 0.55)$

Define a suitable binomial random variable.

$P(X = 50) = \binom{80}{50}(0.55)^{50}(0.45)^{30} = 0.0365$

Use your calculator to find the exact probability using a binomial distribution. ← Statistics 2 Section 1.1

X can be approximated by the normal distribution $Y \sim N(\mu, \sigma^2)$, where $\mu = 80 \times 0.55 = 44$

Use $\mu = np$

$\sigma = \sqrt{80 \times 0.55 \times (1 - 0.55)} = \sqrt{19.8}$ (3 s.f.)

$Y \sim N(44, 19.8)$

Write down the normal approximation.

$P(X = 50) \approx P(49.5 < Y < 50.5)$
$= 0.9280 - 0.8918 = 0.0362$ (4 d.p.)

To estimate the probability that X takes a single value, apply a continuity correction by considering values half a unit below and half a unit above.

$\text{Percentage error} = \frac{0.0365 - 0.0362}{0.0365} \times 100\% = 0.82\%$

Exercise 3B SKILLS REASONING; PROBLEM-SOLVING

1 For each of the following binomial random variables X:
   i state, with reasons, whether X can be approximated by a normal distribution
   ii if appropriate, write down the normal approximation to X in the form $N(\mu, \sigma^2)$, giving the values of μ and σ.

   a $X \sim B(120, 0.6)$ **b** $X \sim B(20, 0.5)$ **c** $X \sim B(250, 0.52)$
   d $X \sim B(300, 0.85)$ **e** $X \sim B(400, 0.48)$ **f** $X \sim B(1000, 0.58)$

2 The random variable $X \sim B(150, 0.45)$. Use a suitable approximation to estimate:
   a $P(X \leqslant 60)$ **b** $P(X > 75)$ **c** $P(65 \leqslant X \leqslant 80)$

3 The random variable $X \sim B(200, 0.53)$. Use a suitable approximation to estimate:
   a $P(X < 90)$ **b** $P(100 \leqslant X < 110)$ **c** $P(X = 105)$

4 The random variable $X \sim B(100, 0.6)$. Use a suitable approximation to estimate:
   a $P(X > 58)$ **b** $P(60 < X \leqslant 72)$ **c** $P(X = 70)$

5 A fair coin is tossed 70 times. Use a suitable approximation to estimate the probability of obtaining more than 45 heads.

6 The probability of the arrow landing on red when a red and white striped wheel is spun is $\frac{50}{101}$.
In one experiment, the wheel is spun 1200 times.
Estimate the probability that the arrow lands on red in at least half of these spins.

(E) **7** **a** Write down two conditions for which the normal distribution may be used as an approximation to the binomial distribution. **(2 marks)**

A company sells orchids of which 45% produce pink flowers.

A random sample of 20 orchids is taken and X produce pink flowers.

b Find $P(X = 10)$. **(1 mark)**

A second random sample of 240 orchids is taken.

c Using a suitable approximation, find the probability that fewer than 110 orchids produce pink flowers. **(3 marks)**

d The probability that at least q orchids produce pink flowers is 0.2. Find q. **(3 marks)**

(E) **8** A drill bit manufacturer claims that 52% of their bits last longer than 40 hours. A random sample of 30 bits is taken and X last longer than 40 hours.

a Find $P(X < 17)$. **(1 mark)**

A second random sample of 600 drill bits is taken.

b Using a suitable approximation, find the probability that between 300 and 350 bits last longer than 40 hours. **(3 marks)**

(E/P) **9** A breakfast cereal has prizes in 56% of the boxes. A random sample of 100 boxes is taken.

a Find the exact value of the probability that exactly 55 boxes contain a prize. **(1 mark)**

b Find the percentage error when using a normal approximation to calculate the probability that exactly 55 boxes contain prizes. **(4 marks)**

3.3 Approximating a Poisson distribution by a normal distribution

If the mean of a Poisson distribution is large then a normal approximation can be used. As for the binomial distribution, we choose the normal distribution to have the same mean as the original Poisson distribution mean and the same variance as the original Poisson distribution variance.

If λ is large:

- **$X \sim \text{Po}(\lambda)$ can be approximated by $Y \sim N(\lambda, \lambda)$**

'The larger the better' is a simplified answer to the question of how large λ should be before a Poisson distribution can be approximated to a normal distribution. In this course, we regard $\lambda > 10$ to be large.

Example 6

The random variable $X \sim \text{Po}(25)$. Use a normal approximation to estimate:

a $P(X > 30)$ **b** $P(18 \leqslant X < 35)$.

$X \sim \text{Po}(25)$

So $Y \sim N(25, 5^2)$

Since λ is quite large we can use a normal approximation with $\mu = \lambda = 25$ and $\sigma^2 = \lambda = 25 = 5^2$

a $P(X > 30) \approx P(Y > 30.5)$ — Apply a continuity correction.

$= P\left(Z > \frac{30.5 - 25}{5}\right)$ — Standardise using $z = \frac{x - \mu}{\sigma}$

$= P(Z > 1.1)$

$= 1 - 0.8643$

$= 0.1357$

b $P(18 \leqslant X < 35) \approx P(17.5 \leqslant Y < 34.5)$

$= P\left(\frac{17.5 - 25}{5} \leqslant Z < \frac{34.5 - 25}{5}\right)$ — Apply continuity correction and then standardise.

$= P(-1.5 \leqslant Z < 1.9)$

$= 0.4713 + 0.4332$

$= 0.9045$

Your calculator may be able to calculate probabilities such as the ones in Example 6 using the original Poisson distribution (part **a** would be 0.13669 and **b** would be 0.90568). However, in S2, if you are asked to use an approximation you will not receive any credit for simply giving these values.

Example 7

A car hire firm has a large fleet of cars that can be hired by the day, and it is found that the fleet suffers breakdowns at the rate of 21 per week. Assuming that breakdowns occur at a constant rate, randomly in time and independently of one another, use a suitable approximation to estimate the probability that in any one week more than 27 breakdowns occur.

Let X represent the number of breakdowns per week.

$X \sim Po(21)$ so $Y \sim N(21, 21)$ — A normal approximation can be used since λ is large.

$P(X > 27) \approx P(Y > 27.5)$ — Apply the continuity correction.

$= P\left(Z > \frac{27.5 - 21}{\sqrt{21}}\right)$ — Standardise

$= P(Z > 1.4184\ldots)$ — The nearest value in the tables is 1.42

$= 1 - 0.9222$

$= 0.0778$

Notation A calculator would give 0.078034, so answers which round to 0.078 are probably acceptable.

Exercise 3C SKILLS PROBLEM-SOLVING

1 The random variable $X \sim \text{Po}(30)$. Use a suitable approximation to estimate:

a $\text{P}(X \leqslant 20)$ **b** $\text{P}(X > 43)$ **c** $\text{P}(25 \leqslant X \leqslant 35)$

2 The random variable $X \sim \text{Po}(45)$. Use a suitable approximation to estimate:

a $\text{P}(X < 40)$ **b** $\text{P}(X \geqslant 50)$ **c** $\text{P}(43 < X \leqslant 52)$

3 The random variable $X \sim \text{Po}(60)$. Use a suitable approximation to estimate:

a $\text{P}(X \leqslant 62)$ **b** $\text{P}(X = 63)$ **c** $\text{P}(55 \leqslant X < 65)$

4 A radioactive object breaks down at a rate of 14 counts per second. Using a normal approximation for a Poisson distribution, determine the probability that in any given second the counts will be:

a 20, 21 or 22 **b** greater than 10 **c** above 12 but less than 16

5 A company hires out boats on a daily basis. The mean number of boats hired per day is 15. Using the normal approximation for a Poisson distribution, find, for a period of 100 days:

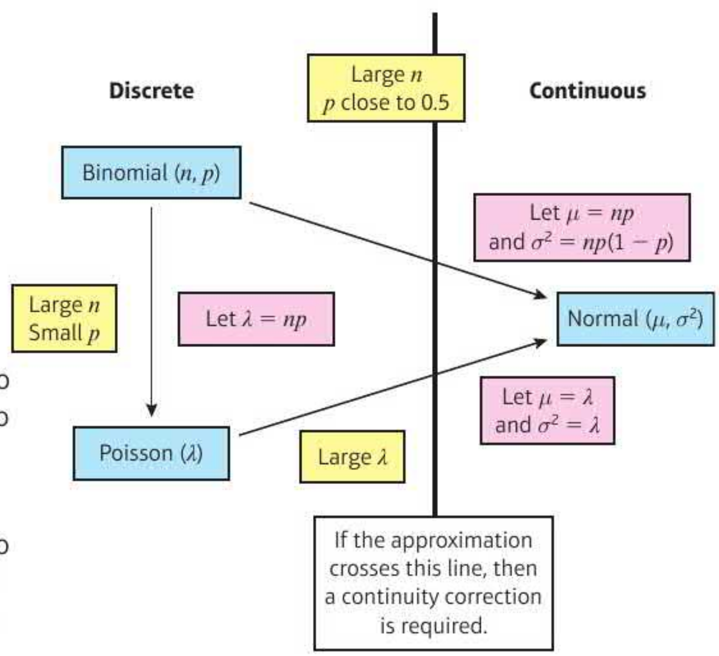
a how often five or fewer boats are hired

b how often exactly 10 boats are hired

c on how many days they will have to turn customers away, if the company owns 20 boats.

3.4 Choosing the appropriate approximation

A binomial distribution can be approximated by a Poisson distribution and sometimes it can be approximated by a normal distribution. The approximations can be summarised by this diagram.



If you are approximating a binomial distribution to a normal distribution, you should always go directly to the normal distribution and not via a Poisson distribution. This is because the binomial to normal approximation involves one not two approximations and should therefore be more accurate.

So, for a binomial distribution there are two possible approximations, depending upon whether p lies close to 0.5 (in which case a normal distribution is used) or p is small (in which case a Poisson distribution is used). If you are unsure about which approximation to use, a helpful guide is to calculate the mean np. If this is less than or equal to 10, you should be able to use the Poisson tables, so that approximation can be used. If the mean is more than 10, then a normal approximation is usually suitable.

Example 8

SKILLS ANALYSIS

A spinner is designed to land on red 10% of the time. Use suitable approximations to estimate the probability of:

a fewer than four reds in 60 spins of the spinner

b more than 20 reds in 150 spins of the spinner.

a Let X represent the number of reds in 60 spins of the spinner.

$X \sim B(60, 0.1)$

$E(X) = 60 \times 0.1 = 6$

so a Poisson approximation can be used.

So $X \sim Po(6)$, approximately

$P(X < 4) = P(X \leqslant 3)$

$= 0.1512$

Calculate the mean of the binomial.

This is justified because $E(X) < 10$

No continuity correction is required. Simply use the Poisson tables with $\lambda = 6$ and $x = 3$.

b Let R represent the number of reds in 150 spins of the spinner.

$R \sim B(150, 0.1)$

$E(R) = 150 \times 0.1 = 15$

so use a normal approximation.

So $X \sim N(15, (\sqrt{13.5})^2)$, approximately

$P(R > 20) \approx P(Y \geqslant 20.5)$

$= P\left(Z > \frac{20.5 - 15}{\sqrt{13.5}}\right)$

$= P(Z > 1.496\ldots)$

$= 1 - 0.9332$

$= 0.0668$

Since the mean is >10 use a normal approximation. The value of $\sigma^2 = 150 \times 0.1 \times 0.9 = 13.5$ and a continuity correction is needed.

Use $z = 1.50$ as this is the nearest value in the tables.

The exact binomial probabilities in Example 8 are **a** 0.13739 and **b** 0.072088

Chapter review 3

(E) **1** The owner of a local corner shop calculates that the probability of a customer buying a newspaper is 0.40.

A random sample of 100 customers is taken.

a Give two reasons why a normal approximation may be used in this situation. **(2 marks)**

b Write down the parameters of the normal distribution used. **(2 marks)**

c Use this approximation to estimate the probability that at least half the customers bought a newspaper. **(2 marks)**

E **2** The random variable $X \sim B(120, 0.46)$

a Find $P(X = 65)$. **(1 mark)**

b State why a normal distribution can be used to approximate X, and write down the parameters of such a normal distribution. **(4 marks)**

c Find the percentage error in using the normal approximation to calculate $P(X = 65)$. **(3 marks)**

E/P **3** The random variable $Y \sim B(300, 0.6)$

a Give two reasons why a normal distribution can be used to approximate Y. **(2 marks)**

b Find, using the normal approximation, $P(150 < Y \leqslant 180)$ **(4 marks)**

c Find the largest value of y such that $P(Y < y) < 0.05$ **(3 marks)**

4 Past records from a supermarket show that 40% of people who buy a bag of nuts choose the large bag. A random sample of 80 people is taken from those who bought bags of nuts. Use a suitable approximation to estimate the probability that more than 30 of these 80 people bought large bags.

E/P **5** A company sells apple-tree seedlings. It is claimed that 55% of these seedlings will produce apples within three years.

A random sample of 20 seedlings is taken and X produce apples within three years.

a Find $P(X > 10)$. **(2 marks)**

A second random sample of 200 seedlings is taken. 95 produce apples within three years.

b Assuming the company's claim is correct, use a suitable approximation to find the probability that 95 or fewer seedlings produce apples within three years. **(4 marks)**

c Using your answer to part **b**, comment on the company's claim. **(1 mark)**

E/P **6** A doctor claims that a specific remedy is successful in curing a particular disease in 52% of cases.

A random sample of 25 people who took the remedy is taken.

a Find the probability that more than 12 people in the sample were cured. **(2 marks)**

A second random sample of 300 people was taken and 170 were cured.

b Assuming the doctor's claim is true, use a suitable approximation to find the probability that at least 170 people were cured. **(4 marks)**

c Using your answer to part **b**, comment on the doctor's claim. **(1 mark)**

7 A fair dice is rolled and the number of sixes obtained is recorded.

Using suitable approximations, find the probability of getting:

a no more than 10 sixes in 48 rolls of the dice

b at least 25 sixes in 120 rolls of the dice.

8 A fair coin is spun 60 times. Use a suitable approximation to estimate the probability of obtaining fewer than 25 heads.

9 The owner of a local shop calculates that the probability of a customer buying a newspaper is 0.40, but the proportion of customers who spend over \$10 is 0.04.

A random sample of 100 customers' shopping is recorded. Use suitable approximations to estimate the probability that in this sample:

a at least half of the customers bought a newspaper

b more than 5 customers spent over \$10.

10 Street light failures in a town occur at a rate of one every two days. Assuming that X, the number of street light failures per week, has a Poisson distribution, find the probability that the number of street lights that will fail in a given week is:

a exactly 2

b less than 6.

Using a suitable approximation, estimate the probability that:

c there will be fewer than 45 street light failures in a 10-week period.

11 Past records from a supermarket show that 20% of people who buy chocolate bars choose the large size. A random sample of 80 people is taken from those who had bought chocolate bars.

a Use a suitable approximation to estimate the probability that more than 20 of these 80 people bought large size bars.

The probability of a customer buying an extra large chocolate bar is 0.02.

b Using a suitable approximation, estimate the probability that fewer than 5 customers in a sample of 150 buy an extra large chocolate bar.

Challenge

A box contains 500 electrical switches. Each has a probability of 0.005 of being faulty, independent of the others. Let X represent the number of defective switches in a box of 500.

a Find $P(X < 2)$.

b Use a Poisson approximation to the binomial distribution and approximate $P(X < 2)$.

c Now use the normal approximation to the binomial distribution to approximate $P(X < 2)$.

d For each approximation, calculate the percentage error and state which is the most accurate approximation.

Summary of key points

1 If X has a binomial distribution with $X \sim B(n, p)$, and

- n is large
- p is small

then X can be approximated by $Po(\lambda)$, where $\lambda = np$.

2 If n is large and p is close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where

- $\mu = np$
- $\sigma = \sqrt{np(1-p)}$

3 If you are using a normal approximation to a binomial distribution, you need to apply a **continuity correction** when calculating probabilities.

4 The random variable $X \sim \text{Po}(\lambda)$ can be approximated by $Y \sim \text{N}(\lambda, \lambda)$ when λ is large (> 10). A continuity correction should be used.

4 CONTINUOUS RANDOM VARIABLES

Learning objectives

After completing this chapter you should be able to:

- Understand and use the probability density function for a continuous random variable → pages 50–55
- Understand and use the cumulative distribution function for a continuous random variable → pages 56–61
- Find the mean, variance, mode, median and percentiles of a continuous random variable → pages 61–75

2.1
2.2
2.3
2.4
2.5

Prior knowledge check

1 The discrete random variable X has probability function:

$$P(X = x) = \begin{cases} k(x-1) & x = 2, 3, 4, 5 \\ \frac{1}{2} & x = 6 \end{cases}$$

where k is a constant. Find:

a the value of k **b** $P(X \geqslant 5)$

c $E(X)$ ← Statistics 1 Sections 6.1, 6.3

2 The random variable Y has $E(Y) = 2$ and $E(Y^2) = 7$. Find:

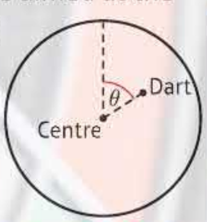
a $E(2Y)$ **b** $Var(Y)$

c $Var(4Y - 2)$ ← Statistics 1 Sections 6.3, 6.4

3 $\int_a^{2a} (3x + 1)\,dx = 168$, where a is a constant.

Find the value of a. ← Pure 2 Section 8.1

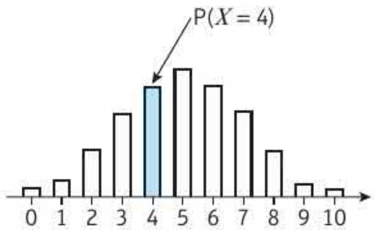
A continuous distribution can be used to model a random variable that is equally likely to take any value in a given range. If a dart is aimed at the centre of a dartboard, the angle θ can be modelled using a continuous distribution.

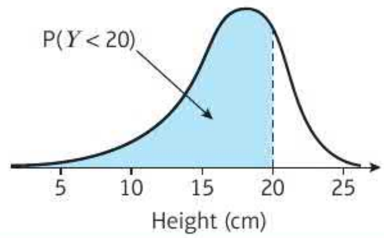


4.1 Continuous random variables

A **continuous random variable** can take any one of **infinitely** many values.
The probability that a continuous random variable takes any one specific value is 0.
However, you can write the probability that it takes values within a given range.

If ten coins are flipped:





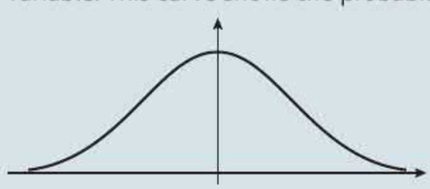
X = number of heads
Probability of getting 4 heads is written as $P(X = 4)$
X is a discrete random variable

Y = maximum height reached by a flipped coin
Probability that the maximum height is less than 20 centimetres (cm) is written as $P(Y < 20)$
Y is a continuous random variable

To describe the probability distribution of a discrete random variable, you usually give a table of values to define the probability of the random variable taking each value in its sample space.

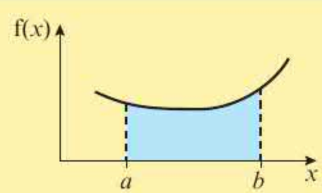
Because the probability of a continuous random variable taking a specific value is 0, you cannot define its distribution in this way. Instead, you can use a **probability density function** (**p.d.f.**) to define the probability of the random variable taking values within a given range.

Links A normally distributed random variable is an example of a continuous random variable. This curve shows the probability density function for a normal distribution:



← **Statistics 1 Section 7.1**

- **If X is a continuous random variable with probability density function $f(x)$, then:**
  - $f(x) \geqslant 0$ for all $x \in \mathbb{R}$ — Probabilities must always be non-negative.
  - $P(a < X < b) = \int_a^b f(x)\,dx$ — This is the area under the p.d.f. between the limits a and b.



  - $\int_{-\infty}^{\infty} f(x)\,dx = 1$ — The total area under the p.d.f. must equal 1.

In practice, probability density functons are often non-zero on a limited subset of the real numbers. For the final condition, you only have to integrate the probability density function across all values of x for which $f(x)$ is non-zero.

Example 1

In each case, state whether the function could be a probability density function, where k is a positive constant.

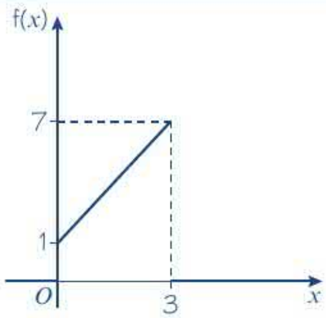
a $f(x) = \begin{cases} 2x + 1 & 0 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$

b $f(x) = \begin{cases} k(2 - x) & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

c $f(x) = \begin{cases} kx(5 - x) & 0 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$

Online Explore probability density functions using GeoGebra.

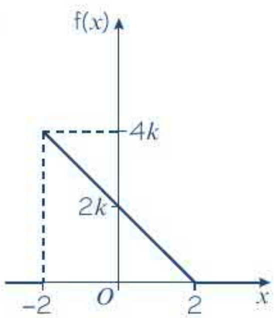
a



Area under $f(x) = 12 \neq 1$, so $f(x)$ cannot be a p.d.f.

Sketch the function. The total area under the function between $x = 0$ and $x = 3$ must equal 1 if it is to be a p.d.f.

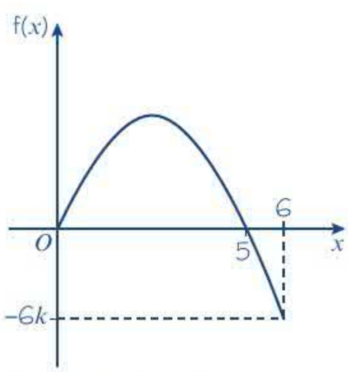
b



Area under $f(x) = 8k$, so $f(x)$ could be a p.d.f. if $k = \frac{1}{8}$

Watch out x can take positive or negative values, but $f(x)$ must be non-negative for all possible values of x.

c



For all positive values of k, there are values in $0 \leqslant x \leqslant 6$ for which $f(x) < 0$, so $f(x)$ cannot be a p.d.f.

The problem states that k is a positive constant.

Example 2 SKILLS INTERPRETATION

The random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(4x - x^2) & 2 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the value of k and sketch the p.d.f.

b $P(2.5 < X < 3)$, giving your answer to 3 decimal places.

a

$$\int_2^4 k(4x - x^2)\,dx = 1$$

$$k\left[2x^2 - \frac{x^3}{3}\right]_2^4 = 1$$

$$k\left(\left(32 - \frac{64}{3}\right) - \left(8 - \frac{8}{3}\right)\right) = 1$$

$$k\left(\frac{16}{3}\right) = 1$$

$$k = \frac{3}{16}$$

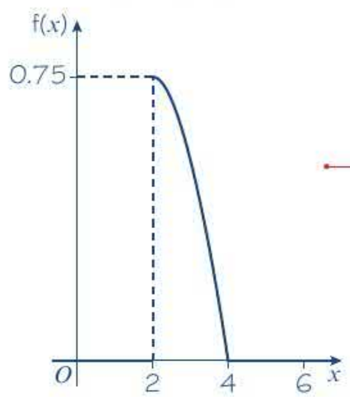
Area under the curve must equal 1.

$\int_2^4 f(x)\,dx = 1$

The p.d.f. is equal to zero everywhere other than [2, 4], so you only need to integrate between these limits.

$\int_{-\infty}^{\infty} f(x)\,dx = \int_2^4 f(x)\,dx$

Sketching the graph:



When sketching the graph, remember to label the axes and label the important values. These are the **boundaries** of the given range of x values and their corresponding y values. Here they are 2 and 4 on the x-axis, and 0.75 on the y-axis.

Watch out $f(x) = 0$ for $x < 2$ and $x > 4$. Make sure that you draw the curve correctly between $x = 2$ and $x = 4$ and draw $f(x) = 0$ elsewhere.

b

$$P(2.5 < X < 3) = \int_{2.5}^{3} \frac{3x(4 - x)}{16}\,dx$$

$$= \frac{3}{16}\int_{2.5}^{3} (4x - x^2)\,dx$$

$$= \frac{3}{16}\left[2x^2 - \frac{1}{3}x^3\right]_{2.5}^{3}$$

$$= \frac{3}{16}\left(9 - \frac{175}{24}\right) = \frac{41}{128}$$

$$= 0.320 \text{ (3 d.p.)}$$

The probability that X takes a value between a and b is given by $\int_a^b f(x)\,dx$. Because the random variable is continuous, it doesn't matter whether you use strict (< or >) or non-strict (≤ or ≥) inequalities.

Watch out Your answer is a probability. Make sure it is between 0 and 1.

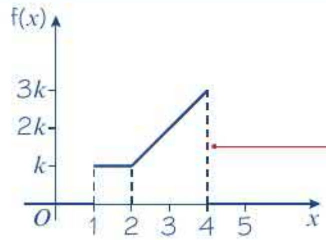
Example 3 SKILLS INTERPRETATION

The random variable X has a probability density function given by:

$$f(x) = \begin{cases} k & 1 < x < 2 \\ k(x-1) & 2 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch f(x). **b** Find the value of k. **c** Find $P(X > 3)$.

a



Use dotted lines to show where the function changes.

b $\int_1^2 k\,dx + \int_2^4 k(x-1)\,dx = 1$

$[kx]_1^2 + \left[\frac{kx^2}{2} - kx\right]_2^4 = 1$

$k + ((8k - 4k) - (2k - 2k)) = 1$

$5k = 1$

$k = \frac{1}{5}$

The total area is 1. Here the area has been found by integrating but it is sometimes easier to find the value of k by looking at the sketch.

Area $= k + \frac{1}{2} \times 2 \times (k + 3k) = 5k$

(using the area of a rectangle and a trapezium)

Area = 1

$5k = 1$

$k = \frac{1}{5}$

c $P(X > 3) = \frac{1}{2}\left(\frac{2}{5} + \frac{3}{5}\right) \times 1 = \frac{1}{2}$

Use the formula for the area of a trapezium.

Exercise 4A SKILLS REASONING

1 Give reasons why the following are not valid probability density functions.

a $f(x) = \begin{cases} \frac{1}{4}x & -1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$ **b** $f(x) = \begin{cases} x^2 & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$ **c** $f(x) = \begin{cases} x^3 - 2 & -1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$

2 For what value of k is the following a valid probability density function?

$$f(x) = \begin{cases} k(x^2 - 1) & -4 \leqslant x \leqslant -2 \\ 0 & \text{otherwise} \end{cases}$$

3 Sketch the following probability density functions.

a $f(x) = \begin{cases} \frac{1}{8}(x-2) & 2 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$ **b** $f(x) = \begin{cases} \frac{2}{15}(5-x) & 1 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$

4 Find the value of k so that each of the following is a valid probability density function.

a $f(x) = \begin{cases} kx & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$ **b** $f(x) = \begin{cases} kx^2 & 0 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$ **c** $f(x) = \begin{cases} k(1 + x^2) & -1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

5 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(4-x) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k.

b Sketch the probability density function for all values of x.

c Find $P(X > 1)$.

E **6** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx^2(2-x) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Find $P(0 < X < 1)$. **(3 marks)**

E **7** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx^3 & 1 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Find $P(1 < X < 2)$ **(3 marks)**

E/P **8** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k & 0 < x < 2 \\ k(2x-3) & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(2 marks)**

b Sketch the probability density function for all values of x. **(2 marks)**

A different continuous random variable Y has a probability density function given by:

$$f(y) = \begin{cases} \frac{3}{16}y^2 & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

c Given that X and Y are independent, find the probability that X and Y are both less than 1. **(4 marks)**

E/P **9** The length of time visitors spent on a news website, X minutes, is modelled using the probability density function:

$$f(x) = \begin{cases} \frac{1}{60}(x+1) & 0 \leqslant x \leqslant 10 \\ 0 & \text{otherwise} \end{cases}$$

a Use this model to find the probability that a randomly chosen visitor spends less than 30 seconds on the website. **(3 marks)**

b Sketch the probability density function. **(2 marks)**

In reality, a small number of visitors are found to spend more than 10 minutes on the website.

c Sketch a probability density function that might provide a better model for X. **(1 mark)**

10 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k & 0 \leqslant x < 1 \\ k(x-1)^2 & 1 \leqslant x < 2 \\ k(3-x) & 2 \leqslant x < 3 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k.

b Find $P(0.5 \leq X \leq 1.5)$

(E) **11** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{k}{x} & 1 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Find $P(2 < X < 4)$, giving your answer in the form $\frac{\ln a}{\ln b}$, where a and b are integers to be determined. **(3 marks)**

(E) **12** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{k}{x+2} & -1 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Find $P(1 < X < 3)$, giving your answer correct to 3 decimal places. **(3 marks)**

(E) **13** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} k\sin(\pi x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Sketch the probability density function for all values of x. **(2 marks)**

c Find $P(0 < X < \frac{1}{3})$. **(3 marks)**

Challenge

The length, T minutes, of a telephone call to a customer service department has probability density function:

$$f(t) = \begin{cases} \frac{k}{t^3} & t \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Given that $\int_1^\infty \frac{1}{t^3}\,dt = \frac{1}{2}$, find the value of k.

b Find the probability that a call will be:

i less than 3 minutes **ii** more than 20 minutes.

c Given that $P(p < T < 2p) = 0.12$, find the value of p.

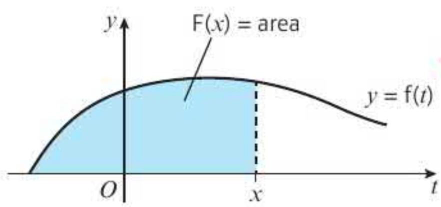
4.2 The cumulative distribution function

Calculating probabilities from the p.d.f. can take a lot of time, and requires integration. You can save time by finding the **cumulative distribution function** (**c.d.f.**) of a random variable.

- For a random variable X, the cumulative distribution function $F(x) = P(X \leqslant x)$.

This is equivalent (equal in value) to the area under the p.d.f. to the left of x.

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt$$



You use t rather than x as the variable in the integration to avoid confusion with x as the limit of the integration.

Notation The p.d.f. is written with a **lower case f**, and the c.d.f. is written with a **capital F**.

Notation When using the cumulative distribution function for the normal distribution in Statistics 1, we have used $\Phi(z)$ to represent $P(Z \leq z)$. Φ (Phi) is the Greek letter equivalent of F.

- If X is a continuous random variable with c.d.f. $F(x)$ and p.d.f. $f(x)$:

$$f(x) = \frac{d}{dx}F(x) \text{ and } F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Notation p.d.f. → Integrate → c.d.f.; c.d.f. → Differentiate → p.d.f.

Example 4

The random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{4}x & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find $F(x)$.

Online Explore cumulative distribution functions using GeoGebra.

Method 1

$$F(x) = \int_{1}^{x} \frac{1}{4}t\,dt$$

$$= \left[\frac{t^2}{8}\right]_1^x$$

$$= \frac{x^2}{8} - \frac{1}{8}$$

Use $F(x) = \int_{-\infty}^{x} f(t)\,dt$. This definition is given in the formulae booklet.

Problem-solving The p.d.f. is 0 for all values of $x < 1$, so you can use 1 as the lower limit for the integration.

Method 2

$$F(x) = \int \frac{1}{4}x\,dx$$

$$= \frac{x^2}{8} + c$$

An alternative method is to use an **indefinite integral** and add $+ c$. You can use $f(x)$ here since there are no limits in the integration.

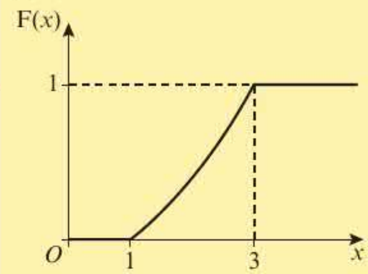
$$\frac{3^2}{8} + c = 1$$

$$c = -\frac{1}{8}$$

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2}{8} - \frac{1}{8} & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

c can be found using F(3) = 1 or F(1) = 0, as 3 and 1 are the upper and lower limits of the given range.

You must define F(x) over all of ℝ. F(x) = 0 for all values less than 1 and F(x) = 1 for all values greater than 3.



Example 5 SKILLS INTERPRETATION

The random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{5} & 1 < x < 2 \\ \frac{1}{5}(x-1) & 2 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Specify fully the cumulative distribution function of X.

Method 1

If $x \leqslant 1$

$F(x) = 0$ so $F(1) = 0$

If $1 < x < 2$

$$F(x) = F(1) + \int_1^x \frac{1}{5}\,dt$$

$$= \left[\frac{1}{5}t\right]_1^x$$

$$= \frac{1}{5}x - \frac{1}{5}$$

so $F(2) = \frac{1}{5}$

If $2 \leqslant x \leqslant 4$

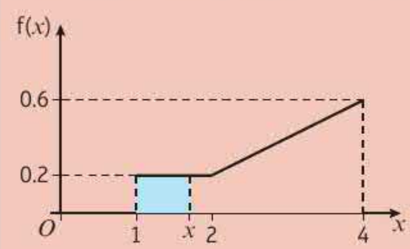
$$F(x) = F(2) + \int_2^x \frac{1}{5}(t-1)\,dt$$

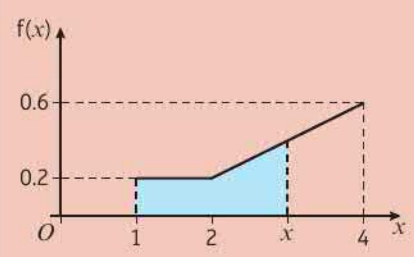
$$= \frac{1}{5} + \left[\frac{t^2}{10} - \frac{t}{5}\right]_2^x$$

$$= \left(\frac{1}{5}\right) + \left(\left(\frac{x^2}{10} - \frac{x}{5}\right) - \left(\frac{4}{10} - \frac{2}{5}\right)\right)$$

$$= \frac{x^2}{10} - \frac{x}{5} + \frac{1}{5}$$

Watch out Integrate each section of the p.d.f. separately. The c.d.f. is **cumulative** so for each section, you need to add on the value of the c.d.f. at the upper limit of the previous section.





Method 2

If $1 < x < 2$

$$F(x) = \int \frac{1}{5}\,dx = \frac{1}{5}x + c$$

$$\frac{1}{5} + c = 0 \text{ so } c = -\frac{1}{5}$$

For the first part, use F(1) = 0.

If $2 \leqslant x \leqslant 4$

$$F(x) = \int \frac{1}{5}(x-1)\,dx = \frac{x^2}{10} - \frac{x}{5} + d$$

$$1 = \frac{4^2}{10} - \frac{4}{5} + d \text{ so } d = \frac{1}{5}$$

For the second part, use F(4) = 1.

$$F(x) = \begin{cases} 0 & x \leqslant 1 \\ \frac{1}{5}x - \frac{1}{5} & 1 < x < 2 \\ \frac{x^2}{10} - \frac{x}{5} + \frac{1}{5} & 2 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$$

Watch out Remember to write the cumulative distribution in full. This means you need to define F(x) for all values of $x \in \mathbb{R}$.

Example 6 SKILLS PROBLEM-SOLVING

The random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{5}x + \frac{3}{20}x^2 & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

Find:

a $P(X \leqslant 1.5)$

b $P(0.5 \leqslant X \leqslant 1.5)$

c $P(X = 1)$

d the probability density function, f(x).

a $P(X \leqslant 1.5) = F(1.5)$

$$= \tfrac{1}{5} \times 1.5 + \tfrac{3}{20} \times 1.5^2$$

$$= 0.6375$$

Using $F(x) = P(X \leqslant x)$

b $P(0.5 \leqslant X \leqslant 1.5) = F(1.5) - F(0.5)$

$$= 0.6375 - 0.1375$$

$$= 0.5$$

$P(0.5 \leqslant X \leqslant 1.5) = P(X \leqslant 1.5) - P(X \leqslant 0.5)$

c $P(X = 1) = 0$

In a continuous distribution, $P(X = x) = 0$

d $\frac{d}{dx}\left(\frac{1}{5}x + \frac{3}{20}x^2\right) = \frac{1}{5} + \frac{3}{10}x$

Use $\frac{d}{dx}F(x) = f(x)$

$$f(x) = \begin{cases} \frac{1}{5} + \frac{3}{10}x & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Watch out If F(x) is constant on a given interval, then f(x) = 0 on that interval. For $x > 2$, F(x) = 1 and $\frac{d}{dx}(1) = 0$.

Exercise 4B

1 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $F(x)$.

2 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{4}(4 - x) & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find $F(x)$.

(E) 3 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{9}x & 0 < x < 3 \\ \frac{1}{9}(6 - x) & 3 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$$

Define fully the cumulative distribution function of X. **(6 marks)**

4 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k & 0 \leqslant x < 3 \\ k(2x - 5) & 3 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch $f(x)$. **b** Find the value of k. **c** Find $F(x)$.

(E) 5 The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{5}(x^2 - 4) & 2 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find the probability density function, $f(x)$. **(3 marks)**

6 The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2}(x - 1) & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find:

a $P(X \leqslant 2.5)$ **b** $P(X > 1.5)$ **c** $P(1.5 \leqslant X \leqslant 2.5)$

E/P **7** The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{6}x^p + q & 2 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$$

Find the exact values of p and q, showing your working clearly. **(7 marks)**

Problem-solving

Use the fact that F(2) = 0 and F(4) = 1 to form two simultaneous equations in p and q.

E **8** The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2}(x^3 - 2x^2 + x) & 1 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

a Find the probability density function $f(x)$. **(3 marks)**

b Sketch the probability density function. **(2 marks)**

c Find $P(X < 1.5)$. **(1 mark)**

E **9** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(4 - x^2) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = \frac{3}{16}$ **(3 marks)**

b Find the cumulative distribution function of X. **(3 marks)**

c Find $P(0.69 < X < 0.70)$. Give your answer correct to 1 significant figure (1 s.f.) **(2 marks)**

E/P **10** The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{120}(kx - x^3) & 0 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

where k is a constant.

Find:

a the value of k **(2 marks)**

b $P(X > 2)$ **(2 marks)**

E **11** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{x \ln 7} & 1 \leqslant x < 7 \\ 0 & \text{otherwise} \end{cases}$$

Find $F(x)$. **(3 marks)**

(E/P) **12** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \pi\cos(\pi x) & 0 \leqslant x < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Find F(x). **(3 marks)**

(E/P) **13** The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 1 \\ k(x - 1 + \ln x) & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find:

a the exact value of k **(3 marks)**

b the probability density function f(x). **(3 marks)**

Challenge

SKILLS INNOVATION; ADAPTIVE LEARNING

The lifetime, in years, of a light bulb is modelled by the random variable T with probability density function:

$$f(t) = \begin{cases} 1.25e^{-1.25t} & t \geqslant 0 \\ 0 & t < 0 \end{cases}$$

Find:

a an expression for F(t)

b the probability that a light bulb lasts between 1 and 2 years

c the probability that a light bulb lasts for more than 3 years.

4.3 Mean and variance of a continuous distribution

You can extend the ideas of mean and variance of a random variable to continuous random variables.

- **If X is a continuous random variable with probability density function f(x):**
  - **the mean or expected value of X is given by:**

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x)\,dx$$

  - **the variance of X is given by:**

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

Links These definitions correspond to the mean and variance of a discrete random variable, with $\sum$ replaced with $\int_{-\infty}^{\infty}$ and the probability function replaced with the p.d.f.

← Statistics 1 Sections 6.3, 6.4

These definitions will be given in the formulae booklet in your exam.

You can also find the mean of a function of a continuous random variable in a similar way as with discrete random variables:

- **If X is a continuous random variable, then $E(g(X)) = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx$**

This gives the following convenient way to calculate Var(X):

- $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$

In the case where g(X) is a linear function of the form $aX + b$, it is useful to learn the following results:

- $\text{E}(aX + b) = a\text{E}(X) + b$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$

Links These results are the same as those for discrete random variables.
← Statistics 1 Sections 6.3, 6.4

SKILLS INNOVATION; ANALYSIS

Challenge

For a continuous random variable, prove the two results above.

Example 7

A random variable Y has probability density function:

$$f(y) = \begin{cases} \frac{1}{4}y & 1 \leqslant y \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a E(Y) **b** Var(Y) **c** E($2Y - 3$) **d** Var($2Y - 3$)

a $\text{E}(Y) = \int_1^3 \frac{1}{4}y^2\,dy$

$= \left[\frac{1}{12}y^3\right]_1^3$

$= \frac{27}{12} - \frac{1}{12} = \frac{26}{12} = \frac{13}{6}$

$yf(y) = y \times \frac{1}{4}y$

If an exact answer is required, you must leave your answer as a fraction. Otherwise you may write the answer as a fraction or as a decimal to 3 significant figures.

b $\text{Var}(Y) = \int_1^3 \frac{1}{4}y^3\,dy - \left(\frac{13}{6}\right)^2$

$= \left[\frac{1}{16}y^4\right]_1^3 - \left(\frac{13}{6}\right)^2$

$= \frac{81}{16} - \frac{1}{16} - \frac{169}{36} = \frac{11}{36}$

$y^2f(y) = y^2 \times \frac{1}{4}y$

c $\text{E}(2Y - 3) = 2\text{E}(Y) - 3$

$= 2 \times \frac{13}{6} - 3$

$= \frac{4}{3}$

$\text{E}(aX + b) = a\text{E}(X) + b$

d $\text{Var}(2Y - 3) = 4\text{Var}(Y)$

$= \frac{44}{36} = \frac{11}{9}$

$\text{Var}(aX + b) = a^2\text{Var}(X)$

Example 8

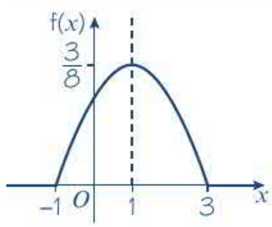
A random variable X has probability density function:

$$f(x) = \begin{cases} \frac{3}{32}(3 + 2x - x^2) & -1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function.

b Find $E(X)$.

a The sketch of the p.d.f. is



b By symmetry, $E(X) = 1$

Problem-solving

If the p.d.f. of a continuous random variable X is **symmetric** about some point $x = a$, then $E(X) = a$.

You could also find $E(X)$ by calculating $\frac{3}{32}\int_{-1}^{3} x(3 + 2x - x^2)\,dx$, but symmetry saves you time in this question.

Example 9

The random variable X has probability density function:

$$f(x) = \begin{cases} \frac{2}{15}x & 0 \leqslant x < 3 \\ \frac{1}{5}(5 - x) & 3 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(X)$ **b** $Var(X)$

a $E(X) = \int_0^3 \frac{2}{15}x^2\,dx + \int_3^5 \frac{1}{5}(5x - x^2)\,dx$

$= \left[\frac{2}{45}x^3\right]_0^3 + \left[\frac{1}{5}\left(\frac{5}{2}x^2 - \frac{x^3}{3}\right)\right]_3^5$

$= \left(\frac{6}{5} - 0\right) + \left(\left(\frac{25}{2} - \frac{25}{3}\right) - \left(\frac{9}{2} - \frac{9}{5}\right)\right)$

$= \frac{8}{3}$

b $Var(X) = \int_0^3 \frac{2}{15}x^3\,dx + \int_3^5 \frac{1}{5}(5x^2 - x^3)\,dx - \left(\frac{8}{3}\right)^2$

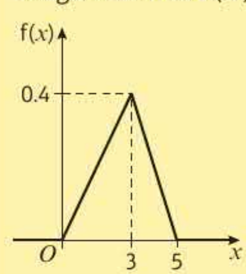
$= \left[\frac{2}{60}x^4\right]_0^3 + \left[\frac{1}{5}\left(\frac{5}{3}x^3 - \frac{x^4}{4}\right)\right]_3^5 - \frac{64}{9}$

$= \left(\frac{27}{10} - 0\right) + \left(\left(\frac{125}{3} - \frac{125}{4}\right) - \left(9 - \frac{81}{20}\right)\right) - \frac{64}{9}$

$= \frac{19}{18}$

To find the **expectation** or variance of a **piecewise function**, you need to integrate each part separately, then add.

Sketching the p.d.f. shows that it is not symmetrical, so you need to integrate to find $E(X)$.



Example 10

SKILLS PROBLEM-SOLVING

In a block-building video game, a cube of side length X pixels is randomly generated.
X is a continuous random variable with probability density function:

$$f(x) = \begin{cases} \frac{1}{800}(x + 10) & 20 \leqslant x \leqslant 40 \\ 0 & \text{otherwise} \end{cases}$$

a Find: **i** $E(X)$ **ii** $E(X^2)$ **iii** $\text{Var}(X)$

b Find the expectation of the volume of the cube.

a i $E(X) = \frac{1}{800}\int_{20}^{40} x(x + 10)\,dx$

$= \frac{1}{800}\left[\frac{1}{3}x^3 + 5x^2\right]_{20}^{40}$

$= \frac{1}{800}\left(\frac{88000}{3} - \frac{14000}{3}\right)$

$= \frac{185}{6}$

ii $E(X^2) = \frac{1}{800}\int_{20}^{40} x^2(x + 10)\,dx$

$= \frac{1}{800}\left[\frac{1}{4}x^4 + \frac{10}{3}x^3\right]_{20}^{40}$

$= \frac{1}{800}\left(\frac{2560000}{3} - \frac{200000}{3}\right)$

$= \frac{2950}{3}$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

iii $\text{Var}(X) = \frac{2950}{3} - \left(\frac{185}{6}\right)^2 = \frac{1175}{36}$

You already know $E(X)$ and $E(X^2)$, so you can work out the variance using the formula
$\text{Var}(X) = E(X^2) - (E(X))^2$

b $E(X^3) = \frac{1}{800}\int_{20}^{40} x^3(x + 10)\,dx$

$= \frac{1}{800}\left[\frac{1}{5}x^5 + \frac{5}{2}x^4\right]_{20}^{40}$

$= \frac{1}{800}(26880000 - 1040000)$

$= 32300$

Watch out The expected volume is $E(X^3)$. This is not the same as $(E(X))^3$. You need to use the rule $E(g(X)) = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx$

Exercise 4C

1 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx^2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a k **b** $E(X)$ **c** $\text{Var}(X)$

2 The continuous random variable Y has a probability density function given by:

$$f(y) = \begin{cases} \frac{y^2}{9} & 0 \leqslant y \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(Y)$ **b** $Var(Y)$ **c** the standard deviation of Y.

3 The continuous random variable Y has a probability density function given by:

$$f(y) = \begin{cases} \frac{y}{8} & 0 \leqslant y \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(Y)$

b $Var(Y)$

c the standard deviation of Y

d $P(Y > \mu)$

e $Var(3Y + 2)$

f $E(Y + 2)$

Hint $\mu = E(Y)$. In part **d**, use your answer from part **a**.

(E) **4** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(1 - x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Find k. **(3 marks)**

b Find $E(X)$. **(3 marks)**

c Show that $Var(X) = \frac{1}{18}$ **(2 marks)**

d Find $P(X > \mu)$. **(3 marks)**

5 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} 12x^2(1 - x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $P(X < 0.5)$ **b** $E(X)$

(E/P) **6** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{8}(1 + x^2) & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X. **(2 marks)**

b Write down $E(X)$ **(1 mark)**

c Show that $\sigma^2 = 0.4$ **(3 marks)**

d Find $P(-\sigma < X < \sigma)$ **(3 marks)**

Problem-solving Use symmetry to answer part **b**.

7 The continuous random variable T has a probability density function given by:

$$f(t) = \begin{cases} kt^3 & 0 \leqslant t \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

a Find k. **b** Show that $E(T)$ is 1.6

Find:

c $E(2T + 3)$ **d** $\text{Var}(T)$ **e** $\text{Var}(2T + 3)$ **f** $P(T < 1)$

8 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{x^2}{27} & 0 \leqslant x < 3 \\ \frac{1}{3} & 3 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Draw a sketch of $f(x)$.

Find:

b $E(X)$ **c** $\text{Var}(X)$ **d** the standard deviation, σ, of X.

E/P **9** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{2}(x - 1) & 1 \leqslant x < 2 \\ \frac{1}{6}(5 - x) & 2 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch $f(x)$. **(2 marks)**

Find:

b $E(X)$ **(5 marks)**

c $\text{Var}(X)$ **(4 marks)**

E **10** Telephone calls arriving at a company are transferred immediately by the receptionist to other people working in the company. The time a call lasts, in minutes, is modelled by a continuous random variable T, having a probability density function given by:

$$f(t) = \begin{cases} kt^2 & 0 \leqslant t \leqslant 10 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = 0.003$ **(3 marks)**

Find:

b $E(T)$ **(3 marks)**

c $\text{Var}(T)$ **(2 marks)**

d the probability of a call lasting between 7 and 9 minutes. **(3 marks)**

e Sketch the probability density function. **(2 marks)**

E/P **11** A continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{3}{4} - \frac{3}{16}x^2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(X)$ **(3 marks)**

b $E(X^2)$ **(3 marks)**

c $Var(X)$ **(2 marks)**

Problem-solving

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

E/P **12** The random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{100} & 0 \leqslant x \leqslant 10 \\ 1 & x > 10 \end{cases}$$

a Find $Var(X)$ **(4 marks)**

b Show that $E(X^3) = 400$ **(3 marks)**

E/P **13** A continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{k}{x} & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the value of k **(3 marks)**

b $E(X)$ **(3 marks)**

c $Var(X)$ **(3 marks)**

E/P **14** A continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{c}{x(3-x)} & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $c = \frac{3}{\ln 4}$ **(3 marks)**

b Calculate the mean and variance of X. **(6 marks)**

Challenge

A continuous random variable X has probability density function:

$$f(x) = \begin{cases} 2x & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(\ln X)$.

4.4 Mode, median, quartiles and percentiles

You need to be able to find the **mode** of a continuous random variable.

- **The mode of a continuous random variable is the value of x for which the p.d.f. is a maximum.**

This is the value of x for which the probability distribution is 'most dense'. A random variable can have more than one modal value, though you will usually only be asked to find the mode in cases where the probability density function has a unique maximum value.

Example 11

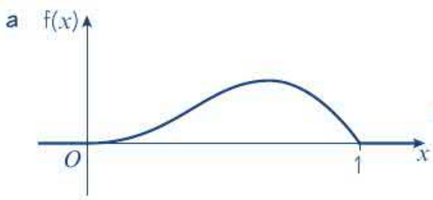
The random variables X and Y have probability density functions $f(x)$ and $g(y)$ respectively.

$$f(x) = \begin{cases} 12x^2(1-x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases} \qquad g(y) = \begin{cases} 2y & 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the mode of:

a X **b** Y

a



Always sketch the graph when finding the mode.

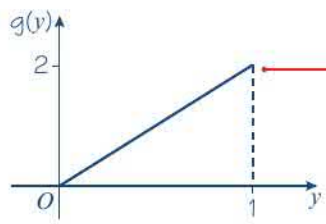
$$\frac{d}{dx}(12x^2 - 12x^3) = 0$$
$$24x - 36x^2 = 0$$
$$12x(2 - 3x) = 0$$
$$x = \tfrac{2}{3}$$

Mode $= \frac{2}{3}$

From the sketch, the mode occurs at the maximum point.

To find the maximum, solve $\frac{d}{dx}f(x) = 0$

b



Mode = 1

You need to justify your answer. You can do this with a sketch, or by observing that $g'(y) = 2 > 0$, so $g(y)$ is strictly increasing on the interval [0, 1], and 0 elsewhere, so the mode must be 1.

Watch out The mode does not need to occur at or even near the 'middle' of a probability distribution.

You can use the cumulative distribution function to define measures of location for a continuous random variable.

- If X is a continuous random variable with c.d.f. $F(x)$:
  - the **median** of X is the value m such that $F(m) = 0.5$
  - the **lower quartile** of X is the value Q_1 such that $F(Q_1) = 0.25$
  - the **upper quartile** of X is the value Q_3 such that $F(Q_3) = 0.75$

Notation The median is also sometimes written as Q_2. For a symmetrical distribution, the median is equal to the mean.

Example 12 SKILLS PROBLEM-SOLVING

A continuous random variable X has probability density function:

$$f(x) = \begin{cases} 4x - 4x^3 & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the c.d.f. of X **b** the median value of X.

a Method 1

$F(x) = \int_0^x (4t - 4t^3)\,dt$ — $F(x) = \int_0^x f(t)\,dt$

$= [2t^2 - t^4]_0^x$

$= 2x^2 - x^4$

Method 2

$F(x) = \int (4x - 4x^3)\,dx$

$= 2x^2 - x^4 + c$

$F(0) = 0$

$c = 0$

$$F(x) = \begin{cases} 0 & x < 0 \\ 2x^2 - x^4 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

b $2m^2 - m^4 = 0.5$ — $F(x) = 0.5$

$2m^4 - 4m^2 + 1 = 0$ — This is a quadratic equation in m^2

$m^2 = \frac{4 \pm \sqrt{16 - 8}}{4}$ — Use the quadratic formula $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$= 1 \pm \frac{\sqrt{2}}{2}$

$m = \sqrt{1 \pm \frac{\sqrt{2}}{2}}$

$= 1.31$ or 0.541 (3 s.f.)

median $= 0.541$ (3 s.f.) — Select the value that is in the range $0 \leqslant x \leqslant 1$

Example 13 SKILLS CRITICAL THINKING

A continuous random variable X has the cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{5}x - \frac{1}{5} & 1 \leqslant x \leqslant 2 \\ \frac{x^2}{10} - \frac{x}{5} + \frac{1}{5} & 2 < x < 4 \\ 1 & x \geqslant 4 \end{cases}$$

Find the interquartile range.

$F(2) = \frac{1}{5}(2) - \frac{1}{5} = 0.2$

Lower quartile

$\frac{(Q_1)^2}{10} - \frac{Q_1}{5} + \frac{1}{5} = 0.25$

$(Q_1)^2 - 2Q_1 + 2 = 2.5$

$(Q_1)^2 - 2Q_1 - 0.5 = 0$

$Q_1 = \frac{2 \pm \sqrt{4 + 2}}{2}$

$= 2.22$ or -0.225 (3 s.f.)

$= 2.22$ (3 s.f.)

Upper quartile

$\frac{(Q_3)^2}{10} - \frac{Q_3}{5} + \frac{1}{5} = 0.75$

$(Q_3)^2 - 2Q_3 + 2 = 7.5$

$(Q_3)^2 - 2Q_3 - 5.5 = 0$

$Q_3 = \frac{2 \pm \sqrt{4 + 22}}{2}$

$= 3.55$ or -1.55 (3 s.f.)

$= 3.55$ (3 s.f.)

Interquartile range $= 3.55 - 2.22$

$= 1.33$ (3 s.f.)

Problem-solving

If a c.d.f. is defined piecewise, it is a good idea to find the **boundary values** for each section of F(x). This will tell you which section of the function to use when calculating the median, quartiles or percentiles.

F(2) = 0.2, so the lower quartile lies in the second section of the function. Set $F(Q_1) = 0.25$ using this section of the c.d.f.

Select the value that is in the range $2 < x < 4$

Select the value that is in the range $2 < x < 4$

Interquartile range $= Q_3 - Q_1$

You can use the cumulative distribution function to define other measures of location for a continuous random variable, such as percentiles.

The nth percentile of P_n is the value a such that $F(a) = \frac{n}{100}$

Example 14 SKILLS REASONING

A continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{10}x & 0 \leqslant x < 2 \\ \frac{1}{4} - \frac{1}{40}x & 2 \leqslant x < 10 \\ 0 & \text{otherwise} \end{cases}$$

a Find:

i the mean of X

ii the mode of X.

b Find the 15th percentile.

a i $\int_0^2 \frac{1}{10}x^2\,dx + \int_2^{10}\left(\frac{1}{4}x - \frac{1}{40}x^2\right)dx$

$= \left[\frac{1}{30}x^3\right]_0^2 + \left[\frac{1}{8}x^2 - \frac{1}{120}x^3\right]_2^{10}$

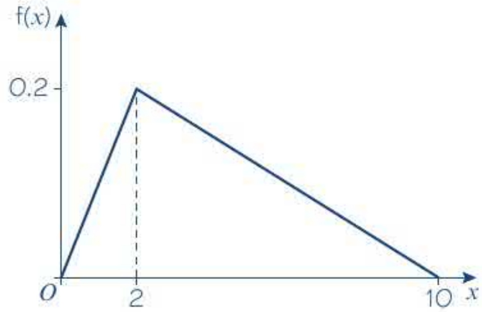
$= \frac{8}{30} + \left(\frac{25}{6} - \frac{13}{30}\right)$

$= 4$

Mean of $X = 4$

When the p.d.f. is defined as a piecewise function, you have to integrate each section of the function separately between the appropriate limits.

ii



From the sketch, the mode of X is 2.

When the sections of a probability density function are linear, it is usually easiest to find the mode by drawing a sketch of the p.d.f. The mode is the value of x at the point where f(x) is a maximum.

b P_{10} is such that $F(a) = 0.1$

Find F(x) and decide in which range P_{10} lies.

$0 \leqslant x < 2$

$F(x) = \int_0^x \frac{1}{10}t\,dt$

$F(x) = \left[\frac{t^2}{20}\right]_0^x = \left(\frac{x^2}{20}\right) - (0)$

Therefore

$F(x) = \frac{x^2}{20}$

Since we are trying to solve F(a) = 0.1, we have found the range that contains P_{10}

and

$F(2) = \frac{4}{20} = 0.2$

$F(a) = 0.1$

$\frac{a^2}{20} = 0.1$

$a^2 = 2$

and so

$a = 1.41$

We ignore the negative root since the answer must be between 0 and 2.

Exercise 4D

SKILLS PROBLEM-SOLVING; REASONING

1 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{80}(8 + 2x - x^2) & 0 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X.

b Find the mode of X.

2 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{8}x & 0 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Find the cumulative distribution function of X.

b Find the median, giving your answer to 3 significant figures.

3 The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{6} & 0 \leqslant x < 2 \\ -\frac{x^2}{3} + 2x - 2 & 2 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find the following, giving your answers to 3 decimal places:

a the median value of X

b the quartiles and the interquartile range of X.

4 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} 1 - \frac{1}{2}x & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X.

b Write down the mode of X.

Find:

c the cumulative distribution function of X

d the median value of X

e the upper quartile.

5 The continuous random variable Y has a probability density function given by:

$$f(y) = \begin{cases} \frac{1}{2} - \frac{1}{9}y & 0 \leqslant y \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of Y.

b Write down the mode of Y.

Find:

c the cumulative distribution function of Y

d the median value of Y correct to 3 significant figures.

6 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{4}x^3 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X.

b Write down the mode of X.

Find:

c the cumulative distribution function of X

d the median value of X.

7 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{8}(x^2 + 1) & -1 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X.

b What can you say about the mode of X?

c Write down the median value of X.

d Find the cumulative distribution function of X.

8 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{10}(3x - x^2) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X.

Find:

b the mode of X

c the cumulative distribution function of X.

SKILLS INNOVATION

d Show that the median value of X lies between 1.23 and 1.24

Hint Consider F(1.23) and F(1.24)

E/P

9 The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{8}(x^2 - 1) & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find:

a the probability density function of the random variable X **(2 marks)**

b the mode of X **(2 marks)**

c the median of X. **(2 marks)**

SKILLS CREATIVITY

d Find the value k such that $P(k < x < k + 1) = 0.6$ **(3 marks)**

E **10** The continuous random variable X has a cumulative distribution function given by:

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0 \\ 4x^3 - 3x^4 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

Find:

a the probability density function of the random variable X **(3 marks)**

b the mode of X **(2 marks)**

c $\mathrm{P}(0.2 < X < 0.5)$ **(3 marks)**

E/P **11** The amount of vegetables eaten by a family in a week is a continuous random variable W kg. The continuous random variable W has a probability density function given by:

$$\mathrm{f}(w) = \begin{cases} \frac{20}{5^5}w^3(5 - w) & 0 \leqslant w \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Find the cumulative distribution function of the random variable W. **(3 marks)**

b Show that the median of W lies between 3.4 kg and 3.5 kg. **(3 marks)**

c Find the mode of W, fully justifying your answer. **(4 marks)**

E/P **12** The continuous random variable X has a probability density function given by:

$$\mathrm{f}(x) = \begin{cases} \frac{1}{4} & 0 \leqslant x < 1 \\ \frac{x^3}{5} & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $\mathrm{E}(X)$ **(5 marks)**

b the cumulative distribution function **(4 marks)**

c to 3 decimal places, the median and the interquartile range of the distribution. **(5 marks)**

d Find, to 3 decimal places, the 40th percentile. **(2 marks)**

E **13** The continuous random variable X has a probability density function given by:

$$\mathrm{f}(x) = \begin{cases} \frac{1}{x \ln 5} & 2 \leqslant x \leqslant 10 \\ 0 & \text{otherwise} \end{cases}$$

a Find the mode of X, fully justifying your answer. **(2 marks)**

b Specify the cumulative distribution function of X. **(3 marks)**

Find:

c the exact value of the median of X **(2 marks)**

d the quartiles and the interquartile range of X. **(3 marks)**

E/P **14** The life, X, of the *Nitelite* light bulb is modelled by the probability density function:

$$f(x) = \begin{cases} 2.5e^{-2.5x} & x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

where X is measured in thousands of hours.

Find:

a the median of X **(5 marks)**

b the quartiles and the interquartile range of X. **(3 marks)**

E/P **15** The continuous random variable X has a probability density function given by:

SKILLS ADAPTIVE LEARNING

$$f(x) = \begin{cases} k\sec^2(\pi x) & 0 \leqslant x \leqslant 0.25 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the value of k **(3 marks)**

b the cumulative distribution function of X **(2 marks)**

c the median of X. **(2 marks)**

E/P **16** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \dfrac{k}{x(5-x)} & 2 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the exact value of k **(4 marks)**

b $E(X)$ **(3 marks)**

c $\text{Var}(X)$ **(4 marks)**

d the c.d.f. of X **(4 marks)**

e the median of X. **(2 marks)**

f Write down the mode of X. **(1 mark)**

g Find, to 3 significant figures, the 80th percentile. **(2 marks)**

Challenge

1 For each of the following sets of conditions, sketch the probability density function of a distribution which satisfies all the conditions:

a the distribution is symmetrical, but mode ≠ median

b there is a unique mode which lies outside the interquartile range.

2 By fully specifying a suitable p.d.f., give an example of a non-symmetrical distribution in which the median and the mode are equal.

Chapter review 4

1 The random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{3}\left(1 + \frac{x}{2}\right) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a $E(X)$ and $E(3X + 2)$

b $\text{Var}(X)$ and $\text{Var}(3X + 2)$

c $P(X < 1)$

d $P(X > E(X))$

e $P(0.5 < X < 1.5)$

2 The random variable X has a probability density function given by:

$$f(x) = \begin{cases} 2 - 2x & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

a Evaluate $E(X)$.

b Evaluate $\text{Var}(X)$.

c Write down the values of $E(2X + 1)$ and $\text{Var}(2X + 1)$.

d Specify fully the cumulative distribution function of X.

e Work out the median value of X.

3 The continuous random variable Y has a cumulative distribution function given by:

$$F(y) = \begin{cases} 0 & y < 1 \\ k(y^2 - y) & 1 \leqslant y \leqslant 2 \\ 1 & y > 2 \end{cases}$$

where k is a positive constant.

a Show that $k = \frac{1}{2}$

b Find $P(Y < 1.5)$.

c Find the value of the median.

d Specify fully the probability density function $f(y)$.

4 The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{5}(x^2 - 4) & 2 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

Find:

a $P(X > 2.4)$

b the median

c the probability density function, $f(x)$.

d Evaluate $E(X)$.

e Find the mode of X.

E **5** The random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx^2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

a Show that $k = \frac{3}{8}$ **(1 mark)**

b Calculate $E(X)$. **(3 marks)**

c Specify fully the cumulative distribution function of X. **(4 marks)**

d Find the value of the median. **(2 marks)**

e Find the value of the mode. **(1 mark)**

E **6** The random variable Y has a probability density function given by:

$$f(y) = \begin{cases} k(y^2 + 2y + 2) & 1 \leqslant y \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

a Show that $k = \frac{3}{62}$ **(2 marks)**

b Specify fully the cumulative distribution function of Y. **(4 marks)**

c Evaluate $P(Y \leqslant 2)$. **(3 marks)**

E **7** A random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{3}{32}(4 - x^2) & -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X. **(2 marks)**

b Write down the mode of X. **(1 mark)**

c Specify fully the cumulative distribution function of X. **(4 marks)**

d Find $P(0.5 < X < 1.5)$. **(3 marks)**

E **8** A random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{3} & 0 \leqslant x < 1 \\ \frac{2}{7}x^2 & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find $E(X)$. **(3 marks)**

b Specify fully the cumulative distribution function of X. **(4 marks)**

c Find:

i the median of X **(3 marks)**

ii the 90th percentile. Give your answer to 3 significant figures. **(1 mark)**

E/P **9** The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.05a^x - b & 1 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

where a and b are positive constants.

Find a and b, showing your working clearly. **(7 marks)**

E/P **10** A student writes the following cumulative distribution function for a continuous random variable X.

$$F(x) = \begin{cases} 0 & x < 5 \\ \frac{1}{5}(16x - x^2 - 55) & 5 \leqslant x \leqslant 10 \\ 1 & x > 10 \end{cases}$$

Explain why this cannot be a cumulative distribution function. **(2 marks)**

E/P **11** A continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} kx - k & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

a Show that $k = \frac{1}{2}$ **(2 marks)**

b Find $E(X)$. **(3 marks)**

c Work out the cumulative distribution function $F(x)$. **(4 marks)**

d Show that the median value lies between 2.4 and 2.5 **(3 marks)**

E/P **12** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} x & 0 \leqslant x \leqslant 1 \\ \dfrac{3x^2}{14} & 1 < x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X. **(2 marks)**

b Find the mode of X. **(1 mark)**

c Find $E(2X)$. **(3 marks)**

d Find $\text{Var}(2X + 1)$. **(3 marks)**

e Specify fully the cumulative distribution function of X. **(4 marks)**

f Using your answer to part **e**, find the median of X. **(2 marks)**

E/P **13** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{x^3}{16} & 0 \leqslant x < 2 \\ \frac{5-x}{6} & 2 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $f(x)$ for all values of x. **(2 marks)**

b Write down the mode of X. **(1 mark)**

c Show that $P(X > 2) = 0.75$ **(2 marks)**

d Define fully the cumulative distribution function $F(x)$. **(4 marks)**

e Find the median of X. **(3 marks)**

E/P **14** The continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{81}(-2x^3 + 15x^2 - 44) & 2 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$

a Find the probability density function $f(x)$. **(3 marks)**

b Find the mode of X. **(2 marks)**

c Sketch $f(x)$ for all values of x. **(3 marks)**

d Find the mean μ of X. **(3 marks)**

e Show that $F(\mu) > 0.5$ **(1 mark)**

f Show that the median of X lies between the mode and the mean. **(3 marks)**

E/P **15** A continuous random variable X has a cumulative distribution function given by:

$$F(x) = \begin{cases} 0 & x < 0 \\ k(35x - 2x^2) & 0 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$

a Show that $k = \frac{1}{125}$ **(1 mark)**

b Find the median of X. **(3 marks)**

c Find the probability density function $f(x)$. **(3 marks)**

d Sketch $f(x)$ for all values of x. **(3 marks)**

e Write down the mode of X. **(1 mark)**

f Find $E(X)$. **(3 marks)**

g Find the 5th percentile, giving your answer to 3 decimal places. **(2 marks)**

(P) **16** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} ax + b & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

If $E(X) = \frac{9}{8}$, find the values of a and b.

(E) **17** A continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} k(x+1)^3 & -1 \leqslant x \leqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive integer.

a Show that $k = 4$ **(3 marks)**

Find:

b $E(X)$ **(4 marks)**

c the cumulative distribution function $F(x)$ **(4 marks)**

d the median. **(3 marks)**

18 The delay in departure, T hours, of a flight from Abu Dhabi International airport is modelled by the probability density function:

$$f(t) = \begin{cases} \frac{1}{72}(6-t)^2 & 0 \leqslant t \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$$

a Find the cumulative distribution function $F(t)$.

b Find the median value of T.

c Find $E(T)$.

(E) **19** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \dfrac{2}{(2x-1)\ln 5} & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Find $F(x)$. **(3 marks)**

(E) **20** *ROBU Bank* sets a test for everybody who applies for a job there. Over the years, it has found that the percentage scored X (measured in 100s) by possible employees can be modelled by the probability density function given by:

$$f(x) = \begin{cases} kx\sin(\pi x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Find:

a the value of k **(5 marks)**

b $E(X)$ **(5 marks)**

(E) **21** A random variable X has a probability density function given by:

$$f(x) = \begin{cases} k & 0 \leqslant x \leqslant 1 \\ \dfrac{k}{x^2} & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of k. **(3 marks)**

b Calculate the value of $\mathrm{E}(X)$. **(4 marks)**

c Calculate the value of $\mathrm{Var}(X)$. **(4 marks)**

Challenge

A continuous random variable X having a probability density function given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

where λ is a positive constant, is said to follow an **exponential** distribution.

Show that:

a $\mathrm{E}(X) = \dfrac{1}{\lambda}$ and $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$

b $\mathrm{P}(X > a + b \mid X > a) = \mathrm{P}(X > b)$

Summary of key points

1 If X is a continuous random variable with **probability density function** $f(x)$, then

- $f(x) \geqslant 0$ for all $x \in \mathbb{R}$
- $\mathrm{P}(a < X < b) = \int_a^b f(x)\,dx$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

2 For a random variable X, the **cumulative distribution function** $F(x) = \mathrm{P}(X \leqslant x)$.

3 If X is a continuous random variable with c.d.f. $F(x)$ and p.d.f. $f(x)$:

$$f(x) = \frac{d}{dx}F(x) \text{ and } F(x) = \int_{-\infty}^{x} f(t)\,dt$$

4 If X is a continuous random variable with probability density function $f(x)$:

- the mean or expected value of X is given by

$$\mathrm{E}(X) = \mu = \int_{-\infty}^{\infty} x f(x)\,dx$$

- the variance of X is given by

$$\mathrm{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

5 If X is a continuous random variable, then $\mathrm{E}(\mathrm{g}(X)) = \int_{-\infty}^{\infty} \mathrm{g}(x)\,\mathrm{f}(x)\,\mathrm{d}x$

6 You can calculate $\mathrm{Var}(X)$ using:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

7 In the case where $\mathrm{g}(X)$ is a linear function of the form $aX + b$,

- $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

8 The mode of a continuous random variable is the value of x for which the p.d.f. is a maximum.

9 If X is a continuous random variable with c.d.f. $\mathrm{F}(x)$:

- the median of X is the value m such that $\mathrm{F}(m) = 0.5$
- the lower quartile of X is the value Q_1 such that $\mathrm{F}(Q_1) = 0.25$
- the upper quartile of X is the value Q_3 such that $\mathrm{F}(Q_3) = 0.75$

Review exercise 1

1 Let $X \sim B(200, 0.02)$

a Write down $E(X)$ and $Var(X)$.

b Suggest why X can be approximated by the Poisson distribution.

c By using a suitable approximation, find $P(X < 6)$.

← Statistics 2 Section 1.3

2 A company manages to respond to $\frac{1}{5}$ of all emails within two hours. The company receives 20 emails in a given day. Let X represent the number of emails responded to within two hours.

a Write down the distribution of the random variable X.

b Find $P(5 < X \leqslant 11)$

The company claims that $\frac{1}{5}$ of all emails are replied to within two hours.

Each day a sample of 20 emails is taken. If there are between 5 and 11 emails inclusive which meet the target, the company is rewarded \$1000. If there are more than 11 emails which meet the target, the company is rewarded with \$2000.

c Calculate the expected reward that the company will receive.

← Statistics 2 Section 1.2

(E) 3 The random variable $X \sim B(15, 0.32)$. Find:

a $P(X = 7)$ b $P(X \leqslant 4)$

c $P(X < 8)$ d $P(X \geqslant 6)$ **(4)**

← Statistics 2 Section 1.1

(E/P) 4 Accidents on a particular motorway occur at an average rate of 1.5 per week.

a Write down a suitable model to represent the number of accidents per week on this motorway. **(1)**

Find the probability that:

b there will be 2 accidents in the same week **(2)**

c there is at least one accident per week for 3 consecutive weeks **(3)**

d there are more than 4 accidents in a two-week period. **(2)**

← Statistics 2 Section 1.1

(E/P) 5 a State two conditions for which a Poisson distribution is a suitable model to use in statistical work. **(2)**

The number of cars passing an observation point in a 10-minute interval is modelled by a Poisson distribution with mean 1.

b Find the probability that in a randomly chosen 60-minute period there will be:

i exactly 4 cars passing the observation point **(2)**

ii at least 5 cars passing the observation point. **(2)**

The number of other vehicles (i.e. other than cars) passing the observation point in a 60-minute interval is modelled by a Poisson distribution with mean 12.

c Find the probability that exactly one vehicle, of any type, passes the observation point in a 10-minute period. **(4)**

← Statistics 2 Sections 2.1, 2.2

(E/P) 6 Two garden machinery firms hire out equipment independently of each other.

Quikmow hire out lawnmowers at a rate of 1.5 mowers per hour.

Easitrim hire out lawnmowers at a rate of 2.2 mowers per hour.

a In a one-hour period, find the probability that each company hires exactly one lawnmower. **(2)**

b In a one-hour period, find the probability that between them, the two companies hire out 4 lawnmowers. **(3)**

c In a three-hour period, find the probability that the total number of lawnmowers hired out by the two companies is less than 12. **(3)**

← Statistics 2 Sections 2.2, 2.3

(E) **7** A manufacturer places toys in cereal boxes. A random sample of 200 cereal boxes is taken, and the number of toys, x, in each box is observed. The data is summarised as follows:

$$\sum x = 290 \quad \sum x^2 = 702$$

a Calculate the mean and the variance of these data. **(2)**

b Explain why the results in part **a** suggest that a Poisson distribution may be a suitable model for the number of toys in each box of cereal. **(1)**

c Use a suitable Poisson distribution to estimate the probability that a randomly chosen box of cereal will contain at least 2 toys. **(3)**

← Statistics 2 Section 2.4

8 A company makes cameras for smartphones. Faults occur at random at a rate of 6 per 1000 cameras. The probability that there are more than 8 faulty cameras from a sample of 1000 cameras is p.

a Find the value of p.

The probability that a sample of 1000 cameras has more than t faults is at least $2p$.

b Find the largest possible value of t.

← Statistics 2 Section 2.2

9 The number of flaws in a given length of rope occur at the rate of 1.2 per metre.

a State the assumptions you need to make to model this situation as a Poisson distribution.

b Find the probability that:

i in a 5-metre length of rope, there are no flaws

ii in a 0.5-metre length of rope, there are two or more flaws.

← Statistics 2 Section 2.2

(E/P) **10** **a** Write down the conditions for which the Poisson distribution may be used as an approximation to the binomial distribution. **(2)**

A receptionist transfers incoming telephone calls to agents who have specialist knowledge to deal with the call. The probability of the caller being connected to the wrong agent is 0.01

b Find the probability that 2 consecutive calls will be connected to the wrong agent. **(1)**

c Find the probability that more than 1 call in 5 consecutive calls are connected to the wrong agent. **(2)**

The call centre receives 1000 calls each day.

d Find the mean and variance of the number of wrongly connected calls. **(2)**

e Use a Poisson approximation to find, to 3 decimal places, the probability that more than 6 calls each day are connected to the wrong agent. **(3)**

← Statistics 2 Sections 2.1, 2.2, 2.3, 2.4

(E) **11** The random variable X has a binomial distribution $X \sim \text{B}(150, 0.02)$

a Find $\text{P}(X = 3)$ **(2)**

A Poisson random variable, $Y \sim \text{Po}(\lambda)$ is used to approximate X.

b Write down the value of λ and justify the use of a Poisson approximation in this instance. **(2)**

← Statistics 2 Sections 1.1, 3.1

(E) **12** In a manufacturing process, 1.5% of the articles produced are faulty. A random sample of 200 articles is selected, and the number of faulty articles X is recorded.

a Write down the distribution of X. **(2)**

b Find $\text{P}(X = 4)$ **(2)**

c Explain why a Poisson distribution could be used as an approximation for X, and write down the parameter for this approximation. **(2)**

d Use your answer to part **c** to find an estimate for $P(X = 4)$, and calculate the percentage error in your estimate. **(3)**

← Statistics 2 Sections 1.1, 3.1

13 For the random variable X, where $X \sim B(200, 0.01)$ use a suitable approximation to find $P(1 < X < 5)$. Justify your choice of approximation.

← Statistics 2 Section 3.2

14 On a particular train line, delayed trains occur at an average rate of seven per day.

By using a suitable approximation, find the probability that fewer than 75 trains are delayed on this line in a 10-day period.

← Statistics 2 Section 3.3

15 A multiple choice paper has 100 questions, each with 5 responses. A student randomly guesses the answers to each of the questions. The paper is quite hard and so the pass mark is 19.

By using a suitable approximation, find the probability that the student will not pass the paper.

← Statistics 2 Section 3.4

(E) **16** A continuous random variable X has probability density function:

$$f(x) = \begin{cases} k(4x - x^3) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive constant.

a Show that $k = \frac{1}{4}$ **(3)**

b Sketch $f(x)$ **(2)**

Find:

c $E(X)$ **(3)**

d the mode of X **(2)**

e the median of X. **(3)**

← Statistics 2 Sections 4.1, 4.3, 4.4

(E) **17** A continuous random variable X has probability density function:

$$f(x) = \begin{cases} kx(x - 2) & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is a positive integer.

a Show that $k = \frac{3}{4}$ **(3)**

b Given that $E(X) = \frac{43}{16}$, find $\text{Var}(X)$. **(4)**

c Find the cumulative distribution function $F(x)$. **(4)**

d Show that the median value of X lies between 2.70 and 2.75 **(3)**

← Statistics 2 Sections 4.1, 4.3, 4.4

(E) **18** Ben attempts to model the continuous random variable Y with the cumulative distribution function:

$$F_1(y) = \begin{cases} 0 & y < 1 \\ 13y - 4y^2 - 9 & 1 \leqslant y \leqslant 2 \\ 1 & y > 2 \end{cases}$$

a Explain what is wrong with Ben's model. **(2)**

Ben adapts his model to use the following cumulative distribution function:

$$F_2(y) = \begin{cases} 0 & y < 1 \\ k(y^4 + y^2 - 2) & 1 \leqslant y \leqslant 2 \\ 1 & y > 2 \end{cases}$$

Using Ben's second model,

b show that $k = \frac{1}{18}$ **(3)**

c find $P(Y > 1.5)$ **(2)**

d specify fully the probability density function $f(y)$. **(3)**

← Statistics 2 Section 4.3

(E) **19** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} 2(x - 2) & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch $f(x)$ for all values of x. **(2)**

b Write down the mode of X. **(1)**

c Given that $E(X) = \frac{8}{3}$, find $\text{Var}(X)$. **(3)**

d Find the median of X. **(3)**

← Statistics 2 Sections 4.1, 4.3, 4.4

E **20** The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{6}x & 0 \leqslant x < 3 \\ 2 - \frac{1}{2}x & 3 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function of X. **(3)**

b Find the mode of X. **(1)**

c Specify fully the cumulative distribution function of X. **(4)**

d Using your answer to part **c**, find the median of X. **(3)**

e Find the 10th to 90th percentile range, giving your answer correct to 3 decimal places. **(4)**

← Statistics 2 Sections 4.1, 4.2, 4.3, 4.4

E **21** The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 0 \\ 2x^2 - x^3 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

a Find $P(X > 0.3)$ **(2)**

b Verify that the median value of X lies between $x = 0.59$ and $x = 0.60$ **(3)**

c Find the probability density function $f(x)$. **(3)**

d Evaluate $E(X)$. **(3)**

e Find the mode of X. **(2)**

← Statistics 2 Sections 4.1, 4.2, 4.3, 4.4

E **22** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} k & 0 \leqslant x \leqslant 2 \\ \dfrac{k}{x} & 2 < x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = \dfrac{1}{2 + \ln 2}$ **(4)**

b Find $E(X)$ **(3)**

← Statistics 2 Sections 4.3, 4.4

23 The continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < 3 \\ \frac{1}{49}(x^2 - 6x + 9) & 3 \leqslant x \leqslant 10 \\ 1 & \text{otherwise} \end{cases}$$

a Find $P(X > 7)$ **(2)**

b Find $P(X > 8 \mid 4 < X < 9)$ **(3)**

c Find $E(X)$ **(4)**

← Statistics 2 Section 4.2

24 The continuous random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{148}(x^2 + 2) & 1 \leqslant x < 5 \\ k(3x - 5) & 5 \leqslant x \leqslant 7 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = \frac{1}{39}$ **(4)**

b Find the cumulative distribution function. **(6)**

c Find the median. **(3)**

d Find the 80th percentile. **(2)**

← Statistics 2 Section 4.4

Challenge

A continuous random variable X has probability density function:

$$f(x) = \begin{cases} ke^{-x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = 1$.

b Find the cumulative distribution function $F(X)$.

c Hence, or otherwise, find the exact value of $P(1 < X < 4)$.

← Statistics 2 Section 4.2

5 CONTINUOUS UNIFORM DISTRIBUTION

3.1

Learning objectives

After completing this chapter you should be able to:

- Understand, use and model situations using the continuous uniform distribution → pages 88–98

Prior knowledge check

1 The discrete random variable X has probability function:

$$P(X = x) = \begin{cases} k(x-1) & x = 2, 3, 4, 5 \\ \frac{1}{2} & x = 6 \end{cases}$$

where k is a constant. Find:

a the value of k **b** $P(X \geqslant 5)$

c $E(X)$ ← Statistics 1 Sections 6.1, 6.2, 6.3

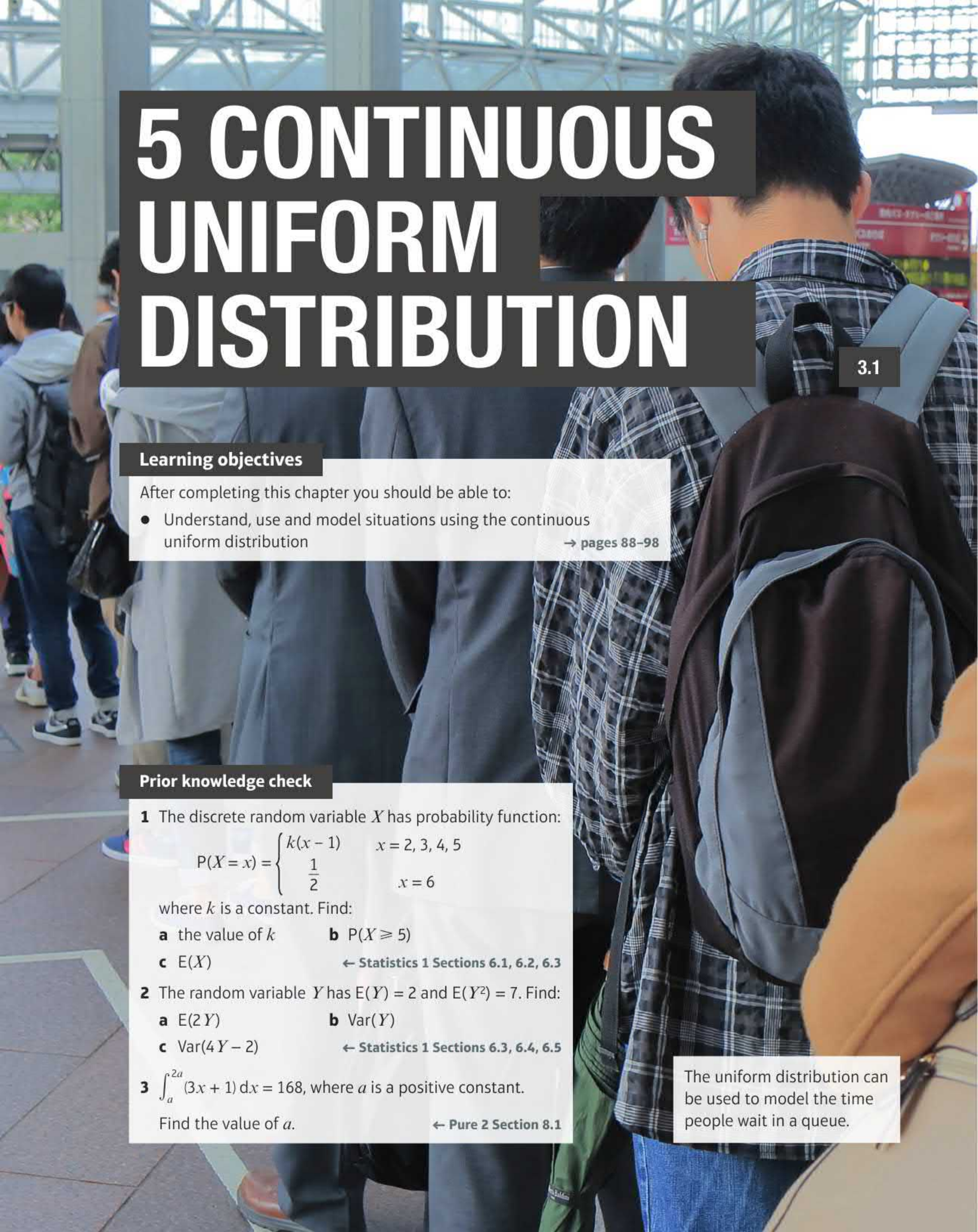
2 The random variable Y has $E(Y) = 2$ and $E(Y^2) = 7$. Find:

a $E(2Y)$ **b** $Var(Y)$

c $Var(4Y - 2)$ ← Statistics 1 Sections 6.3, 6.4, 6.5

3 $\int_a^{2a} (3x + 1)\,dx = 168$, where a is a positive constant.

Find the value of a. ← Pure 2 Section 8.1



The uniform distribution can be used to model the time people wait in a queue.

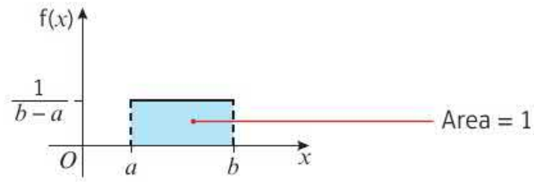
5.1 The continuous uniform distribution

- A random variable having a **continuous uniform distribution** over the interval $[a, b]$ has p.d.f.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

A sketch of the p.d.f. is shown.

Notation If X has the continuous uniform distribution over the interval $[a, b]$, you write $X \sim U[a, b]$.

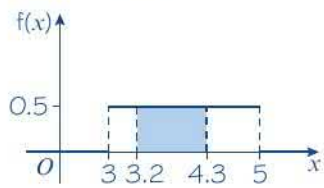


Example 1

The continuous random variable X is **uniformly** distributed on [3, 5]. Find:

a $P(3.2 < X < 4.3)$ **b** k such that $P(2X < k - X) = 0.2$

a $\frac{1}{b-a} = \frac{1}{5-3} = 0.5$



$P(3.2 < X < 4.3) = (4.3 - 3.2) \times 0.5$
$= 0.55$

b $2X < k - X$
$X < \frac{k}{3}$
So $P\left(X < \frac{k}{3}\right) = 0.2$
$0.5\left(\frac{k}{3} - 3\right) = 0.2$
$k = 10.2$

Notation When dealing with a uniform distribution, it is easier to sketch the p.d.f. and work out the area of the rectangle. You can also use proportion: you are interested in a range of values of width 1.1 out of an interval of width 2, so the probability is $\frac{1.1}{2} = 0.55$

$P(3.2 < X < 4.3)$ is the area of the shaded section on the sketch.

Solve the inequality.

Write an expression for the area under the rectangle to the left of $\frac{k}{3}$. Set this equal to 0.2 and then solve to find k.

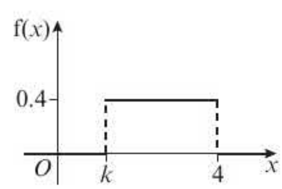
Example 2

SKILLS ADAPTIVE LEARNING

The continuous random variable X has p.d.f. as shown in the diagram. Find:

a the value of k **b** $P(3 < X < 3.5)$

c $P(X > 3 \mid X > 2)$



a Area = 1
$0.4 \times (4 - k) = 1$
$4 - k = 2.5$
$k = 1.5$

b $P(3 < X < 3.5) = 0.4 \times (3.5 - 3) = 0.2$

c $P(X > 3 \mid X > 2) = \dfrac{P(X > 2 \cap X > 3)}{P(X > 2)}$
$= \dfrac{P(X > 3)}{P(X > 2)}$
$= \dfrac{0.4}{0.8} = \dfrac{1}{2}$

Use $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$. The probability that X is greater than 2 **and** greater than 3 is just the probability that X is greater than 3.

← Statistics 1 Section 4.7

Problem-solving

To solve **conditional probability** problems with a continuous uniform distribution, you can use a continuous uniform distribution on a restricted sample space. Given that $X > 2$, the value of X is uniformly distributed on [2, 4].

Example 3

The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{5} & 3 \leqslant x \leqslant 8 \\ 0 & \text{otherwise} \end{cases}$$

a Write down the name of this distribution.

The continuous random variable $Y = 12 - 3X$.

Find:

b $E(Y)$ **c** $P(Y > 0)$ **d** Find $P(X < 7 \mid Y < 0)$

a Continuous uniform distribution

b $E(Y) = E(12 - 3X)$
$= 12 - 3E(X)$
$= 12 - 3 \times 5.5$
$= -4.5$

Use $E(aX + b) = aE(X) + b$

c $P(Y > 0) = P(12 - 3X > 0)$
$= P(12 > 3X)$
$= P(4 > X)$
$= \dfrac{1}{5}$

Convert the probability in terms of Y into a probability in terms of X.

Use proportion. $X < 4$ is an interval of width 1 out of a total interval width of 5.

d $P(X < 7 \mid Y < 0) = P(X < 7 \mid X > 4)$
$= \dfrac{P(4 < X < 7)}{P(X > 4)}$
$= \dfrac{0.6}{0.8}$
$= \dfrac{3}{4}$

Write both probabilities in terms of the same random variable.

The probability that $X < 7$ **and** $X > 4$ is $P(4 < X < 7)$

Problem-solving

You could also tackle this problem by finding the distribution of Y. A linear transformation of a uniform distribution will be uniform, so $Y \sim U[-12, 3]$.

Example 4 SKILLS PROBLEM-SOLVING

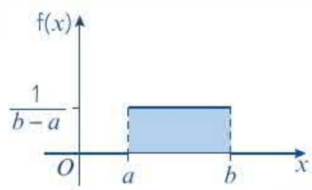
The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{otherwise} \end{cases}$$

Find:

a E(X) **b** Var(X) **c** F(x)

a



By symmetry $E(X) = \dfrac{a+b}{2}$

> The formula for E(X) for the continuous uniform distribution is given in the formulae booklet. However, you should be able to derive it from first principles.

b $\text{Var}(X) = \displaystyle\int_a^b \frac{(x-\mu)^2}{b-a}\,dx$

> This is $\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx$. You could use $E(X^2) - (E(X))^2$ but it is more difficult in this case. You will need to know how to factorise $b^3 - a^3$

$$= \int_a^b \left(x - \left(\frac{a+b}{2}\right)\right)^2 \times \frac{1}{(b-a)}\,dx$$

$$= \left[\frac{\left(x - \left(\frac{a+b}{2}\right)\right)^3}{3(b-a)}\right]_a^b$$

> Using $\int (x+a)^n\,dx = \dfrac{(x+a)^{n+1}}{n+1}$

$$= \frac{\left(b - \left(\frac{a+b}{2}\right)\right)^3}{3(b-a)} - \frac{\left(a - \left(\frac{a+b}{2}\right)\right)^3}{3(b-a)}$$

$$= \frac{\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3}{3(b-a)}$$

$$= \frac{\left(\frac{b-a}{2}\right)^3 + \left(\frac{b-a}{2}\right)^3}{3(b-a)}$$

> Using $-(a-b)^3 = (b-a)^3$

$$= \frac{\frac{(b-a)^3}{8} + \frac{(b-a)^3}{8}}{3(b-a)}$$

$$= \frac{\frac{(b-a)^3}{4}}{3(b-a)}$$

$$= \frac{(b-a)^3}{12(b-a)}$$

$$= \frac{(b-a)^2}{12}$$

> The formula for Var(X) for the continuous uniform distribution is given in the formulae booklet.

c If $a \leqslant x \leqslant b$, $F(x) = \int_a^x \frac{1}{b-a}\,dt$

$$= \left[\frac{t}{b-a}\right]_a^x$$

$$= \frac{x-a}{b-a}$$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & x > b \end{cases}$$

Watch out The cumulative distribution function for a uniform continuous distribution is not given in the formulae booklet. It can be useful to remember it, but make sure you know how to derive it from first principles, as shown here.

- **For a continuous uniform distribution U[a, b]:**
  - $E(X) = \frac{a+b}{2}$
  - $Var(X) = \frac{(b-a)^2}{12}$
  - $F(X) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & x > b \end{cases}$

Example 5

The continuous random variable X is uniformly distributed over the interval [4, 7]. Find:

a $E(X)$

b $Var(X)$

c the cumulative distribution function of X, for all x.

a $E(X) = \frac{4+7}{2} = 5.5$

b $Var(X) = \frac{(7-4)^2}{12} = \frac{3}{4}$

c $F(x) = \int_4^x \frac{1}{7-4}\,dt = \left[\frac{t}{3}\right]_4^x = \frac{x-4}{3}$

$$F(x) = \begin{cases} 0 & x < 4 \\ \frac{x-4}{3} & 4 \leqslant x \leqslant 7 \\ 1 & x > 7 \end{cases}$$

You can write this down straight away if you learn the formula for $F(X)$ of a continuous uniform distribution.

Example 6 SKILLS PROBLEM-SOLVING

The continuous random variable Y is uniformly distributed over the interval $[a, b]$.
Given that $E(Y) = 1$ and $Var(Y) = \frac{16}{3}$, find the value of a and the value of b.

$E(Y) = \frac{a + b}{2} = 1$

$a + b = 2$ **(1)**

$Var(Y) = \frac{(b - a)^2}{12} = \frac{16}{3}$

$(b - a)^2 = 64$ **(2)**

Solving equations **(1)** and **(2)** simultaneously:

$b = 2 - a$

$(2 - a - a)^2 = 64$

$(2 - 2a) = \pm 8$

$2 - 2a = 8$, $a = -3$, $b = 2 - (-3) = 5$

$2 - 2a = -8$, $a = 5$, $b = 2 - 5 = -3$

Since $a < b$, $a = -3$ and $b = 5$.

Problem-solving

Use the formulae for the mean and variance of a continuous uniform distribution to form simultaneous equations in a and b.

Example 7 SKILLS INTERPRETATION

The continuous variable X is uniformly distributed over the interval $[-3, 5]$.

a Write down $E(X)$.

b Use integration to find the variance of X.

a $E(X) = \frac{-3 + 5}{2} = 1$

b $Var(X) = E(X^2) - (E(X))^2$

$= \int_{-3}^{5} \left(\frac{x^2}{8}\right) dx - 1^2$

$= \left[\frac{x^3}{24}\right]_{-3}^{5} - 1$

$= \frac{125}{24} + \frac{27}{24} - 1$

$= \frac{16}{3}$

You could have used

$$\int_{-3}^{5} \frac{(x - 1)^2}{8} dx = \left[\frac{(x - 1)^3}{24}\right]_{-3}^{5}$$

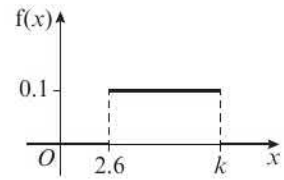
$$= \frac{4^3}{24} - \frac{(-4)^3}{24}$$

$$= \frac{16}{3}$$

Exercise 5A

1 The continuous random variable $X \sim U[2, 7]$.
Find:
a $P(3 < X < 5)$ b $P(X > 4)$

2 The continuous random variable X has a probability density function as shown in the diagram.
Find:
a the value of k b $P(4 < X < 7.9)$



P 3 The continuous random variable X has probability density function:

$$f(x) = \begin{cases} k & -2 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$$

Find:
a the value of k b $P(-1.3 < X < 4.2)$ c p, such that $P(3X < X + p) = 0.5$
d $P(X > 5 \mid X > 0)$ e $P(X > 0 \mid X < 3)$ f $P(X < 1 \mid 0 < X < 2)$

P 4 The continuous random variable $Y \sim U[a, b]$. Given that $P(Y < 5) = \frac{1}{4}$ and $P(Y > 7) = \frac{1}{2}$, find the value of a and the value of b.

P 5 The continuous random variable $X \sim U[2, 8]$.
a Write down the distribution of $Y = 2X + 5$.
b Find $P(12 < Y < 20)$.

Hint If X has a continuous uniform distribution, then $aX + b$, where a and b are constants, will also have a continuous uniform distribution.

E/P 6 The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{10} & 2 \leqslant x \leqslant 12 \\ 0 & \text{otherwise} \end{cases}$$

a Write down the name of this distribution. **(1 mark)**
The continuous random variable $Y = 20 - 2X$.
Find:
b $E(Y)$ **(2 marks)**
c $P(Y < 4)$ **(2 marks)**
d $P(Y > 4 \mid X < 10)$ **(3 marks)**

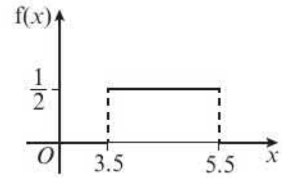
E/P 7 The continuous random variable Y is uniformly distributed over the interval $[-3, 5]$.
Find:
a $E(X)$ **(1 mark)**
b $\text{Var}(X)$ **(1 mark)**
c $E(X^2)$ **(2 marks)**
d the cumulative distribution function of X, for all x. **(3 marks)**

8 Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ for the following probability density functions.

a $f(x) = \begin{cases} \frac{1}{4} & 1 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$

b $f(x) = \begin{cases} \frac{1}{8} & -2 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$

9 The continuous random variable X has probability density function as shown in the diagram.



Find:

a $\mathrm{E}(X)$ **b** $\mathrm{Var}(X)$

c $\mathrm{E}(X^2)$ **d** the cumulative distribution function of X, for all x.

E/P **10** The continuous random variable $Y \sim \mathrm{U}[a, b]$. Given that $\mathrm{E}(Y) = 1$ and $\mathrm{Var}(Y) = \frac{4}{3}$, find the value of a and the value of b. **(3 marks)**

E/P **11** The continuous random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{6} & -1 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

Given that $Y = 4X - 6$, find $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$. **(3 marks)**

E/P **12** The continuous random variable R is uniformly distributed over the interval $\alpha \leqslant R \leqslant \beta$. Given that $\mathrm{E}(R) = 5$ and $\mathrm{Var}(R) = \frac{4}{3}$, find:

a the value of α and the value of β **(3 marks)**

b $\mathrm{P}(R < 5.2)$ **(1 mark)**

E/P **13** The continuous random variable X is uniformly distributed over the interval $\alpha \leqslant R \leqslant \beta$.

a Write down the probability density function of X, for all x. **(1 mark)**

b Given that $\mathrm{E}(X) = 2.5$ and $\mathrm{P}(X < 1) = \frac{4}{11}$, find the value of α and the value of β. **(3 marks)**

E **14** The continuous random variable X is uniformly distributed over the interval $[-5, 4]$.

a Write down fully the probability density function $f(x)$ of X. **(2 marks)**

b Sketch the probability density function $f(x)$ of X. **(2 marks)**

Find:

c $\mathrm{E}(X^2)$ **(2 marks)**

d $\mathrm{P}(-0.2 < X < 0.6)$ **(2 marks)**

E **15** A continuous random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -3 \\ \dfrac{x+3}{7} & -3 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$$

a Find $P(X < 0)$. **(1 mark)**

b Find the probability density function $f(x)$ of X. **(2 marks)**

c Write down the name of the distribution of X. **(1 mark)**

d Find the mean and the variance of X. **(3 marks)**

E/P **16** The continuous random variable X is uniformly distributed over the interval $[-1, 4]$.
Find:

a $E(X)$ **(2 marks)**

b $\text{Var}(X)$ **(2 marks)**

c $E(X^2)$ **(2 marks)**

d $P(X < 1.4)$ **(1 mark)**

A total of 6 observations of X are made.

e Find the probability that exactly 4 of these observations are less than 1.4. **(2 marks)**

E/P **17** The continuous random variable X is uniformly distributed over the interval $[\alpha, \beta]$.
Given that $E(X) = 7.5$ and $P(X > 10.5) = 0.25$,

a find the value of α and the value of β. **(3 marks)**

b Given that $P(X < c) = \frac{1}{3}$, find:

i the value of c

ii $P(c < X < 9)$ **(3 marks)**

Challenge

The continuous random variable X is uniformly distributed over the interval $[\alpha, \beta]$.
Find:

a $P(X < \frac{3}{7}\alpha + \frac{4}{7}\beta)$ **b** $P(X > \frac{2}{5}\alpha + \frac{3}{5}\beta)$

5.2 Modelling with the continuous uniform distribution

The continuous uniform distribution is frequently used to model real-life situations. For example, if you know that trains leave from a station hourly, but you arrive not knowing when the next train will leave, then the length of time you have to wait after arriving at the station, X minutes, could be modelled as $X \sim U[0, 60]$.

Example 8 SKILLS ADAPTIVE LEARNING

The trunk of a small tree varies in diameter from 10 cm at the bottom to 2 cm at the top.
The tree is cut horizontally at a randomly chosen point, and the radius R cm of the cross-section is modelled as $R \sim U[1, 5]$.

Find the expected value of the area, A, of the cross-section of the tree.

$A = \pi R^2$

$E(A) = E(\pi R^2)$

$= \pi E(R^2)$

$Var(R) = E(R^2) - (E(R))^2$

Rearranging gives

$E(R^2) = Var(R) + (E(R))^2$

$Var(R) = \frac{(5-1)^2}{12} = \frac{4}{3}$

$E(R) = \frac{5+1}{2} = 3$

$E(R^2) = \frac{4}{3} + 9 = \frac{31}{3}$

$E(A) = \frac{31\pi}{3}$

Watch out $E(R^2)$ is not the same as $(E(R))^2$

To find $E(R^2)$, you could have used

$E(X^2) = \int x^2 f(x)\,dx$

$E(R^2) = \int_1^5 \frac{1}{4}r^2 dr = \left[\frac{1}{12}r^3\right]_1^5$ since $\frac{1}{b-a} = \frac{1}{5-1} = \frac{1}{4}$

$= \frac{125}{12} - \frac{1}{12}$

$= \frac{31}{3}$

Example 9

SKILLS INTERPRETATION

The length of a pencil is measured to the nearest cm. Write down the distribution of the rounding error, R. (This is the error between the true value and the rounded value.)

The error is the difference between the true length and the recorded length.

If a pencil is recorded as 20 cm long, then its length is anywhere in the interval

$19.5\text{ cm} \leqslant \text{length} < 20.5\text{ cm}$

The error is therefore in the interval

$-0.5 \leqslant \text{error} < 0.5$

As it is reasonable to assume that the error is equally likely to take any of the values in this range

$R \sim U[-0.5, 0.5]$

Notation The uniform distribution is often used as a model for errors made by rounding up or down when recording measurements.

Example 10

SKILLS INTERPRETATION

Write down the name of the distribution you would recommend as a suitable model for each of the following situations.

a The masses of 200 g tins of tomatoes produced on a production line.

b The difference between the true length and the length of metal rods as measured to the nearest cm.

a Normal — You expect more tins to be near the 200 g mark.

b Continuous uniform — It is reasonable to assume that the difference is equally likely to take any of the values in the range −0.5 to 0.5

Exercise 5B

SKILLS INTERPRETATION; PROBLEM-SOLVING

1 The random variable X represents the side length of a square and is modelled as $X \sim \text{U}[4.5, 5.5]$. The random variable Y represents the area of the square.
Find $\text{E}(Y)$.

2 The random variable R has a continuous uniform distribution over the interval [5, 11].

a Specify fully the probability density function of R.

b Find $\text{P}(7 < R < 10)$.

The random variable A represents the area of a circle with radius R cm.

c Find $\text{E}(A)$.

(E) 3 In a computer game, a cat appears every 2 seconds. The player stops the cat by pressing a key. To play, you need to stop the cat as soon as it appears. Given that the player actually presses the key T seconds after the cat first appears, a simple model of the game assumes that T is a continuous uniform random variable defined over the interval [0, 1].

a Write down $\text{P}(T < 0.2)$. **(1 mark)**

b Write down $\text{E}(T)$. **(2 marks)**

c Use integration to find $\text{Var}(T)$. **(3 marks)**

(E/P) 4 The time in minutes that Priya takes to check out at her local supermarket follows a continuous uniform distribution defined over the interval [2, 10].
Find:

a the probability that Priya will take more than 7 minutes to check out on one visit to the supermarket **(2 marks)**

b the probability that Priya will take less than 5 minutes to check out on each of three successive visits to the supermarket. **(3 marks)**

Given that Priya has already spent 5 minutes at the checkout,

c find the probability that she will take a total of less than 8 minutes to check out. **(2 marks)**

(E/P) 5 A drinks machine pours coffee into cups. It is electronically controlled to cut off the flow of coffee randomly between 175 ml and 215 ml. The random variable X represents the volume of coffee poured into a cup, and is uniformly distributed.

a Specify the probability density function of X and sketch its graph. **(3 marks)**

b Find the probability that the machine pours:

i less than 187 ml

ii exactly 187 ml. **(3 marks)**

c Calculate the interquartile range of X. **(3 marks)**

d Determine the value of x such that $\text{P}(X \geqslant x) = 0.65$ **(2 marks)**

SKILLS CRITICAL THINKING

Rashmi buys five cups of coffee from the drinks machine for people in her office.

Hint Use a binomial distribution.

e Find the probability that exactly three of the cups contain less than 187 ml. **(3 marks)**

E/P **6** The continuous random variable X represents the error, in millimetres (mm), made when a machine cuts iron rods to a target length. X has a continuous uniform distribution over the interval $[-3.0, 3.0]$

Find:

a $P(X < -2.3)$ **(2 marks)**

b $P(|X| > 2.0)$ **(2 marks)**

Ten rods are cut.

c Calculate the probability that exactly six are cut within 2 mm of the target length. **(3 marks)**

E/P **7** A manufacturer produces sweets of length Y mm, where Y has a continuous uniform distribution with range [20, 28].

a Find the probability that a randomly selected sweet has length greater than 26 mm. **(2 marks)**

These sweets are randomly packed in bags of 20 sweets.

b Find the probability that a randomly selected bag will contain at least 7 sweets with length greater than 26 mm. **(3 marks)**

E/P **8** The waiting times, in minutes, between flight take-offs at an airport are modelled by the continuous random variable X with probability density function:

$$f(x) = \begin{cases} \frac{1}{5} & 2 \leqslant x \leqslant 7 \\ 0 & \text{otherwise} \end{cases}$$

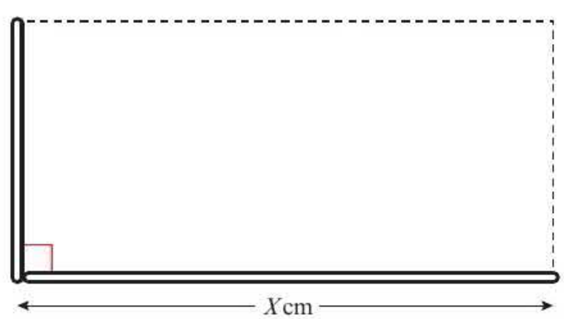
A randomly selected flight takes off at 10 am.

a Find the probability that the next flight takes off before 10:05 am. **(2 marks)**

b Find the probability that at least 3 of the next 10 flights have a waiting time of more than 6 minutes. **(3 marks)**

E/P **9** A wooden rod of length 20 cm is cut into two pieces at a randomly chosen point. The length of the longer piece, X cm, is modelled as having a continuous uniform distribution over the interval [10, 20].

The two pieces of the rod are used to form the base and height of a rectangle, as shown below.



Find the expected area of the rectangle. **(6 marks)**

Chapter review 5

1 The continuous random variable X is uniformly distributed over the interval $[-2, 5]$.

a Sketch the probability density function $f(x)$ of X.

Find:

b $E(X)$

c $Var(X)$

d the cumulative distribution function of X, for all x

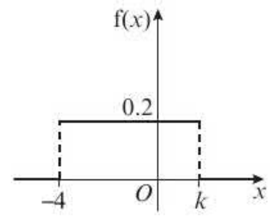
e $P(3.5 < X < 5.5)$

f $P(X = 4)$

g $P(X > 0 | X < 2)$

h $P(X > 3 | X > 0)$

2 The continuous random variable X has p.d.f. as shown in the diagram.



Find:

a the value of k

b $P(-2 < X < -1)$

c $E(X)$

d $Var(X)$

e the cumulative distribution function of X, for all x.

3 The continuous random variable Y is uniformly distributed over the interval $a \leqslant Y \leqslant b$. Given $E(Y) = 2$ and $Var(Y) = 3$, find:

a the value of a and the value of b

b $P(Y > 1.8)$

E/P 4 The continuous random variable X has a continuous uniform distribution over the interval $[0, 2]$. The continuous random variable $Y = 10 - 5X$.

a Describe the distribution of Y. **(2 marks)**

b Find $P(Y < 3)$ **(2 marks)**

c Find $P(Y > 3 | X > 0.5)$ **(3 marks)**

E/P 5 Kwan has a pair of scissors and a piece of string 20 cm long. The string has a mark on one end. Kwan cuts the string into two pieces at a randomly chosen point. Let X represent the length of the piece of string with the mark on it.

a Write down the name of the probability distribution of X and sketch the graph of its probability density function. **(3 marks)**

b Find the values of $E(X)$ and $Var(X)$. **(4 marks)**

c Using your model, calculate the probability that the shorter piece of string is at least 8 cm long. **(3 marks)**

E/P 6 Joan records the temperature every day. The highest temperature she recorded was 29 °C to the nearest degree. Let X represent the error in the measured temperature.

a Suggest a suitable model for the distribution of X. **(1 mark)**

b Using your model, calculate the probability that the error will be less than 0.2 °C. **(3 marks)**

c Find the variance of the error in the measured temperature. **(2 marks)**

(E/P) **7** Jameil catches a bus to work every morning. According to the timetable, the bus is due at 9 am, but Jameil knows that the bus can arrive at a random time between three minutes early and ten minutes late. The random variable X represents the time, in minutes, after 9 am when the bus arrives.

a Suggest a suitable model for the distribution of X and specify it fully. **(2 marks)**

b Calculate the mean value of X. **(2 marks)**

c Find the cumulative distribution function of X. **(4 marks)**

Jameil will be late for work if the bus arrives after 9:05 am.

d Find the probability that Jameil is late for work. **(2 marks)**

(E/P) **8** A plumber measures, to the nearest cm, the lengths of pipes.

a Suggest a suitable model to represent the difference between the true lengths and the measured lengths. **(1 mark)**

b Find the probability that, for a randomly chosen pipe, the measured length will be within 0.2 cm of the true length. **(2 marks)**

c Three pipes are selected at random. Find the probability that the measured lengths of all three pipes will be within 0.2 cm of the true length. **(2 marks)**

(E/P) **9** A coffee machine pours coffee into cups. It is electronically controlled to cut off the flow of coffee randomly between 190 ml and 210 ml. The random variable X represents the volume of coffee poured into a cup.

a Specify the probability density function of X and sketch its graph. **(3 marks)**

b Find the probability that the machine pours:

i less than 198 ml

ii exactly 198 ml. **(3 marks)**

c Calculate the interquartile range of X. **(2 marks)**

d Given that the machine has already poured 195 ml of coffee into a cup, find the probability that it will pour more than 200 ml into that cup. **(2 marks)**

10 Write down the name of the distribution you would recommend as a suitable model for each of the following situations:

a the difference between the true height and the height measured, to the nearest cm, of randomly chosen people

b the heights of randomly selected 18-year-old females.

(E/P) **11** A continuous random variable X is uniform over the interval $[b, 5b]$

a Write down the probability density function of X. **(2 marks)**

b Write down the value of $\mathrm{E}(X)$. **(1 mark)**

c Show by integration that $\mathrm{Var}(X) = \frac{4b^2}{3}$ **(3 marks)**

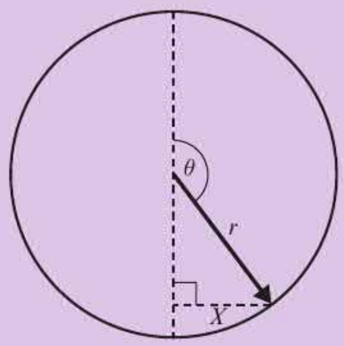
d Given that $b = 3$, find $\mathrm{P}(X > 10)$. **(2 marks)**

Five observations are taken from this distribution.

e Find the probability that exactly three of them are bigger than 10. **(4 marks)**

Challenge

A spinner is made using a circle of radius r, and a pointer of length r which is pivoted at the centre of the circle. The pointer is spun and allowed to come to rest. The random variable θ represents the angle between the vertical and the resting position of the spinner, and the random variable X represents the horizontal distance of the tip of the spinner from the vertical line.



a Describe a suitable distribution to model θ.

b Hence, or otherwise, find $\mathrm{E}(X)$.

c Briefly explain how this spinner could be used as part of an experiment to estimate the value of π.

Summary of key points

1 A random variable having a continuous uniform distribution $\mathrm{U}[a, b]$ has p.d.f.

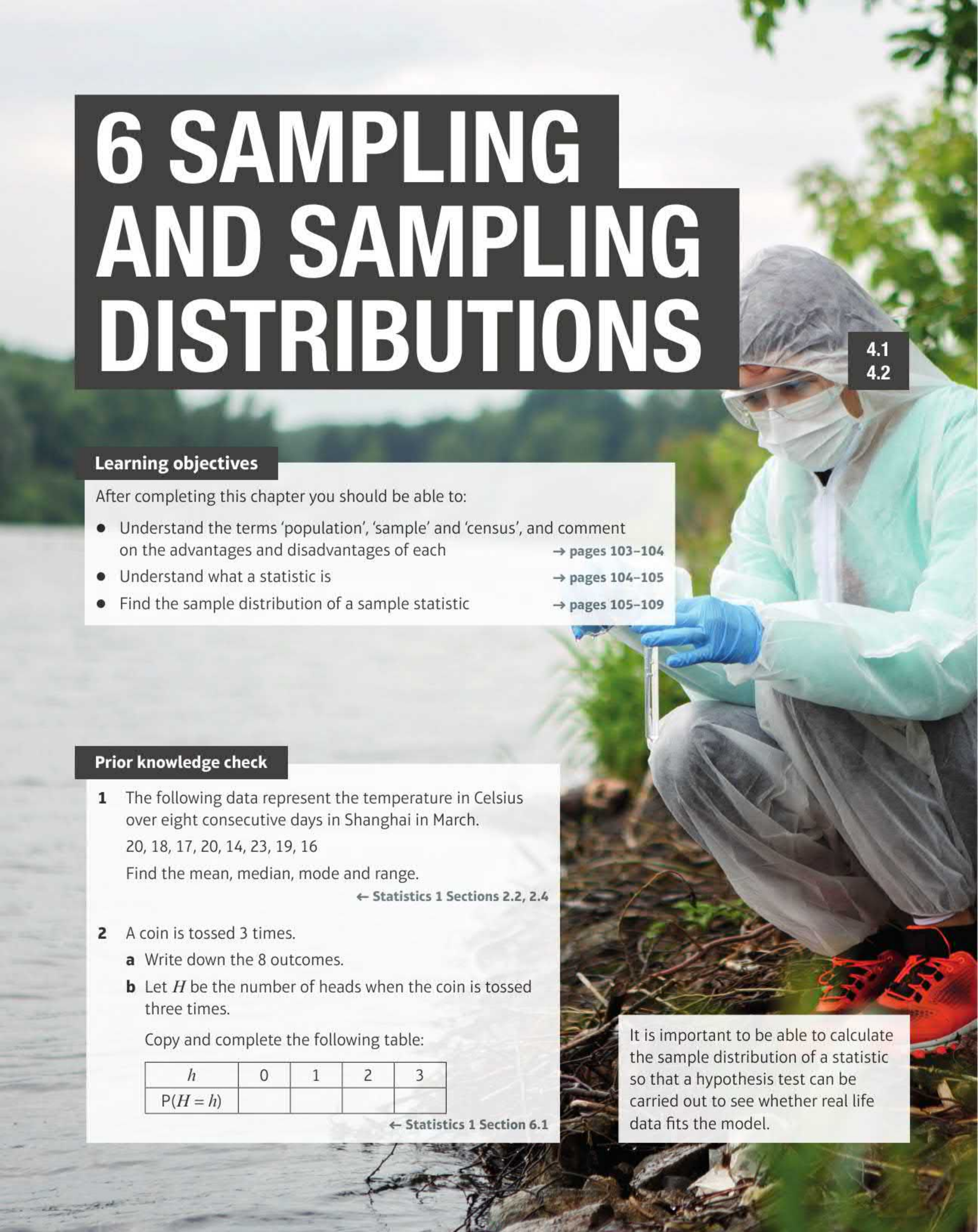
$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{otherwise} \end{cases}$$

2 For a continuous uniform distribution $\mathrm{U}[a, b]$:

$$\mathrm{E}(X) = \frac{a+b}{2}$$

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

$$\mathrm{F}(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leqslant x \leqslant b \\ 1 & x > b \end{cases}$$



6 SAMPLING AND SAMPLING DISTRIBUTIONS

4.1
4.2

Learning objectives

After completing this chapter you should be able to:

- Understand the terms 'population', 'sample' and 'census', and comment on the advantages and disadvantages of each → pages 103–104
- Understand what a statistic is → pages 104–105
- Find the sample distribution of a sample statistic → pages 105–109

Prior knowledge check

1 The following data represent the temperature in Celsius over eight consecutive days in Shanghai in March.

20, 18, 17, 20, 14, 23, 19, 16

Find the mean, median, mode and range.

← Statistics 1 Sections 2.2, 2.4

2 A coin is tossed 3 times.

a Write down the 8 outcomes.

b Let H be the number of heads when the coin is tossed three times.

Copy and complete the following table:

h	0	1	2	3
$\mathrm{P}(H = h)$				

← Statistics 1 Section 6.1

It is important to be able to calculate the sample distribution of a statistic so that a hypothesis test can be carried out to see whether real life data fits the model.

6.1 Populations and samples

- **In statistics, a population is the whole set of items that are of interest.**

For example, the population could be the items manufactured by a factory or all the people in a town.
Information can be obtained from a population. This is known as raw data.

- **A census observes or measures every member of a population.**
- **A sample is a selection of observations taken from a subset of the population which is used to find out information about the population as a whole.**

There are a number of advantages and disadvantages of both a census and a sample.

	Advantages	Disadvantages
Census	• It should give a completely accurate result	• Time-consuming and expensive • Cannot be used when the testing process destroys the item • Hard to process a large quantity of data
Sample	• Less time-consuming and expensive than a census • Fewer people needed to respond • Less data to process than in a census	• The data may not be as accurate • The sample may not be large enough to give information about small sub-groups of the population

The size of a sample can affect the validity of any conclusions drawn.

- The size of the sample depends on the required accuracy and available resources.
- Generally, the larger the sample, the more accurate it is, but you will need greater resources.
- If the population is very varied, you need a larger sample than if the population were uniform.
- Different samples can lead to different conclusions due to the natural variation in a population.

- **Individual units of a population are known as sampling units.**
- **Often, sampling units of a population are individually named or numbered to form a list called a sampling frame.**

Example 1 SKILLS REASONING

A supermarket wants to test a delivery of plums for ripeness by cutting them in half.

a Suggest a reason why the supermarket should not test all the plums in the delivery.

The supermarket tests a sample of 5 plums and finds that 4 of them are ripe.
They estimate that 80% of the plums in the delivery are ripe.

b Suggest one way that the supermarket could improve their estimate.

a Testing all the plums would mean that there were none left to sell.

When testing a product destroys it, a census is not appropriate.

b Testing a larger sample (30 plums, for example) would give a better estimate of the overall proportion of ripe plums.

In general, larger samples produce more accurate predictions about a population.

Exercise 6A SKILLS REASONING

1 A school uses a census to investigate the dietary requirements of its students.

a Explain what is meant by a census.

b Give one advantage and one disadvantage to the school of using a census.

2 A factory makes safety ropes for climbers and has an order to supply 3000 ropes. The buyer wants to know if the load at which the ropes break is more than a certain figure.

a Suggest a reason why a census would not be used for this purpose.

The factory tests four ropes and the load for breaking is recorded for each rope:

320 kg 260 kg 240 kg 180 kg

b The factory claims that the ropes are safe for loads up to 250 kg. Use the sample data to comment on this claim.

c Suggest one way in which the company can improve their prediction.

3 A city council wants to know what people think about its recycling centre. The council decides to carry out a sample survey to learn the opinion of residents.

a Write down one reason why the council should not take a census.

b Suggest a suitable sampling frame.

c Identify the sampling units.

(P) **4** A manufacturer of microswitches is testing how reliable its switches are. It uses a special machine to switch them on and off repeatedly until they break.

a Give one reason why the manufacturer should use a sample rather than a census.

The company tests a sample of five switches and obtains the following results:

23 150 25 071 19 480 22 921 7455

b The company claims that its switches can be operated an average of 20 000 times without breaking. Use the sample data above to comment on this claim.

c Suggest one way the company could improve its prediction.

(P) **5** The manager of a garage wants to know what their mechanics think about a new savings scheme designed for them. The manager decides to ask all the mechanics in the garage.

a Describe the population the manager will use.

b Write down the main advantage in asking all of their mechanics.

6.2 The concept of a statistic

Any characteristic of a population which is measurable is called a **population parameter**. A parameter is a numerical property of a sample. For example, the population mean and the population variance are population parameters.

Notation We usually use Greek letters for population parameters.

Usually the population is too large to calculate these parameters. In order to estimate a population parameter, we take a random sample from the population and use observations from the items in it to estimate the required parameters. For example, we could use the sample mean as an estimate for the popultion mean

If we drew the numbers 3, 8, 32, 38, 43 and 44 from a bag, then the sample mean $\bar{x} = \frac{\Sigma X_i}{n} = \frac{3 + 8 + 32 + 38 + 43 + 44}{6} = 28$ could be an estimate for the parameter μ.

Notation We use upper case letters to represent a distribution. We use X_i to denote an observation from the distribution X.

- **A statistic is a quantity calculated only from the observations in a sample. It does not involve any unknown parameters. Thus, a statistic is a numerical property of a sample.**

The sample mean $\bar{X} = \frac{\Sigma X_i}{n}$ and sample variance $S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n - 1}$ are statistics, but $\frac{\Sigma(X_i - \mu)^2}{n - 1}$ is not, because it involves the population parameter μ.

Example 2 SKILLS REASONING

a Define a statistic.

A random sample $X_1, X_2, ..., X_n$ is taken from a population with unknown mean μ.

b For each of the following, state whether or not it is a statistic.

i $\frac{X_2 + X_5 + X_8}{3}$ **ii** $\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2$ **iii** $\frac{\Sigma X}{n} - \mu^2$

a A quantity calculated only from the observations in a sample.

b **i** and **ii** are statistics but **iii** is not (it involves μ which is unknown).

6.3 The sampling distribution of a statistic

If we repeatedly take samples from a population and calculate the same statistic each time, there is a range of values that the statistic can take. The statistic will have its own distribution which we call the **sampling distribution**.

- **The sampling distribution of a statistic gives all the values of a statistic and the probability that each would happen by chance alone.**

Example 3

A school would like to introduce a school uniform, and thus wants to find out how much support this idea has among the students at the school. The random variable X is defined as:

$$X = \begin{cases} 1 & \text{if the student supports the idea} \\ 0 & \text{otherwise} \end{cases}$$

a Suggest a suitable population and the parameter of interest.

In a random sample, 15 students are asked if they would support the idea.
The random sample is represented by $X_1, X_2, ..., X_{15}$.

b Write down the sampling distribution of the statistic $Y = \sum_{i=1}^{15} X_i$

a The population is the responses of all the school students. In terms of the random variable X, it will consist of 1's or 0's. The parameter of interest is p, the proportion of the population who support the idea.

b ΣY = the number of students in the sample who support the idea. Since the sample is random:

- each observation is independent
- p is constant
- the responses will be either success (1) or failure (0)

These are the conditions for a binomial distribution.
$Y \sim B(15, p)$

Example 4

A manufacturer of light bulbs sells 60 watt and 100 watt bulbs in the ratio of 3 : 1

a Find the mean and variance of the wattage of the light bulbs in this population.

A random sample of 3 light bulbs is taken from a store containing bulbs in this ratio.

b List all the possible samples.

c Find the sampling distribution of the mean $\overline{X}$.

d Find the sampling distribution of the mode M.

a The distribution of the population is:

x	60	100
$P(X = x)$	$\frac{3}{4}$	$\frac{1}{4}$

$\mu = \sum x P(X = x) = 60 \times \frac{3}{4} + 100 \times \frac{1}{4} = 70$ watts

$\sigma^2 = \sum x^2 P(X = x) - \mu^2$

$= 60^2 \times \frac{3}{4} + 100^2 \times \frac{1}{4} - (70)^2$

$= 300$ watts^2

Notation We use lowercase letters to say what values the distribution could take.

b The possible samples are:

(60, 60, 60) (100, 60, 60) (60, 100, 60)
(60, 60, 100) (100, 100, 60) (100, 60, 100)
(60, 100, 100) (100, 100, 100)

Here we have written out all the possible samples of size 3.

c $P(\overline{X} = 60) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$

$P(\overline{X} = 73\frac{1}{3}) = 3 \times \frac{1}{4} \times \left(\frac{3}{4}\right)^2 = \frac{27}{64}$

$P(\overline{X} = 86\frac{2}{3}) = 3 \times \left(\frac{1}{4}\right)^2 \times \frac{3}{4} = \frac{9}{64}$

$P(\overline{X} = 100) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}$

The distribution of $\overline{x}$ is:

$\overline{x}$	60	$73\frac{1}{3}$	$86\frac{2}{3}$	100
$P(\overline{X} = \overline{x})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

P(X = 60) is constant at $\frac{3}{4}$

i.e the case (60, 60, 60)

i.e. cases (100, 60, 60) (60, 100, 60) (60, 60, 100)

i.e. cases (100, 100, 60) (100, 60, 100) (60, 100, 100)

i.e. the case (100, 100, 100)

This shows the distribution of the statistic X by writing down all the possible values and the probability with which each is likely to occur.

d The mode M can only take the values 60 and 100.

$P(M = 60) = \frac{27}{64} + \frac{27}{64} = \frac{27}{32}$

$P(M = 100) = \frac{9}{64} + \frac{1}{64} = \frac{5}{32}$

The distribution of M is:

m	60	100
$P(M = m)$	$\frac{27}{32}$	$\frac{5}{32}$

i.e. cases (60, 60, 60) (100, 60, 60) (60, 100, 60) (60, 60, 100)

i.e. the other cases.

Example 5

A large bag of counters has 10% of the counters labelled with the number 0, 40% with the number 1, and 50% with the number 2. A random sample of 3 counters is taken from the bag.

a List all possible samples.

b Find the sampling distribution of the median N.

Watch out Make sure you list all outcomes; for example, (2, 2, 1) is not the same as (2, 1, 2).

a The possible samples are:

$(0, \underline{0}, 0)$ $(0, \underline{0}, 1)$ $(0, 1, \underline{0})$ $(1, \underline{0}, 0)$ $(0, \underline{0}, 2)$ $(0, 2, \underline{0})$ $(2, \underline{0}, 0)$

$(1, \underline{1}, 0)$ $(1, 0, \underline{1})$ $(0, \underline{1}, 1)$ $(1, \underline{1}, 2)$ $(1, 2, \underline{1})$ $(2, \underline{1}, 1)$ $(1, \underline{1}, 1)$

$(0, \underline{1}, 2)$ $(0, 2, \underline{1})$ $(\underline{1}, 0, 2)$ $(\underline{1}, 2, 0)$ $(2, \underline{1}, 0)$ $(2, 0, \underline{1})$

$(2, \underline{2}, 0)$ $(2, 0, \underline{2})$ $(0, 2, \underline{2})$ $(\underline{2}, 2, 1)$ $(2, 1, \underline{2})$ $(1, \underline{2}, 2)$ $(2, \underline{2}, 2)$

Median = 0 in these

Median = 1 in these

Median = 2 in these

b The median can only take the values 0, 1 and 2.

Let $P(0) = p = \frac{1}{10}$, $P(1) = q = \frac{4}{10}$ and $P(2) = r = \frac{5}{10}$

$$\begin{aligned}P(N = 0) &= P(0,0,0) + P(0,0,1) + P(0,1,0) + P(1,0,0) + P(0,0,2) \\ &\quad + P(0,2,0) + P(2,0,0) \\ &= ppp + ppq + pqp + qpp + ppr + prp + rpp \\ &= \left(\frac{1\times1\times1}{1000}\right) + \left(\frac{1\times1\times4}{1000}\right) + \left(\frac{1\times4\times1}{1000}\right) + \left(\frac{4\times1\times1}{1000}\right) \\ &\quad + \left(\frac{1\times1\times5}{1000}\right) + \left(\frac{1\times5\times1}{1000}\right) + \left(\frac{5\times1\times1}{1000}\right) = \frac{28}{1000} = \frac{7}{250}\end{aligned}$$

$$\begin{aligned}P(N = 1) &= P(1,1,0) + P(1,0,1) + P(0,1,1) + P(1,1,2) + P(1,2,1)\\&\quad + P(2,1,1) + P(1,1,1) + P(0,1,2) + P(0,2,1) + P(1,0,2)\\&\quad + P(1,2,0) + P(2,1,0) + P(2,0,1)\\&= qqp + qpq + pqq + qqr + qrq + rqq + qqq + pqr\\&\quad + prq + qpr + qrp + rqp + rpq\end{aligned}$$

$$= \left(\frac{4\times4\times1}{1000}\right) + \left(\frac{4\times1\times4}{1000}\right) + \left(\frac{1\times4\times4}{1000}\right) + \left(\frac{4\times4\times5}{1000}\right) + \left(\frac{4\times5\times4}{1000}\right) + \left(\frac{5\times4\times4}{1000}\right) + \left(\frac{4\times4\times4}{1000}\right) + \left(\frac{1\times4\times5}{1000}\right) + \left(\frac{1\times5\times4}{1000}\right) + \left(\frac{4\times1\times5}{1000}\right) + \left(\frac{4\times5\times1}{1000}\right) + \left(\frac{5\times4\times1}{1000}\right) + \left(\frac{5\times1\times4}{1000}\right) = \frac{59}{125}$$

$$P(N = 2) = 1 - \left(\frac{7}{250} + \frac{118}{250}\right) = \frac{125}{250} = \frac{1}{2}$$

The distribution of N is:

n	0	1	2
$P(N = n)$	$\frac{7}{250}$	$\frac{59}{125}$	$\frac{1}{2}$

Exercise 6B

1 A forester wants to estimate the height of the trees in a forest. He measures the heights of 50 randomly selected trees and works out the mean height. State with a reason whether or not this mean is a statistic.

2 A random sample $M_1, M_2, M_3, \ldots, M_n$ is taken from a population with unknown mean μ. For each of the following, state whether or not it is a statistic.

i $\dfrac{M_3 + M_8}{2}$ **ii** $\dfrac{\sum M}{n}$ **iii** $\dfrac{\sum M}{n} - \mu^2$

3 The owners of a chain of hairdressing shops want to introduce the use of aprons in all the shops. The random variable Y is defined as:

$Y = 0$ if the staff are happy to wear an apron, and

$Y = 1$ if the staff are unhappy about wearing an apron.

a Suggest a suitable population and identify any parameter of interest.

In a random sample of 20 hairdressers, each one is asked if they are happy or unhappy about wearing an apron.

b Write down the name of the sampling distribution of the statistic $X = \sum_{1}^{20} Y$

4 A receptionist makes spelling mistakes at the rate of 5 mistakes for every 10 pages. He has just finished typing a six-page document.

a Write down a suitable sampling distribution for the number of spelling mistakes in his document.

b Find the probability that there were fewer than 2 spelling mistakes in the document.

Ⓟ **5** A drawer contains a large number of coins. 50% are fifty-cent coins. 25% are twenty-cent coins. 25% are ten-cent coins.

a Find the mean, μ, and the variance, σ, for the value of this population of coins.

A random sample of two coins is chosen from the drawer.

b List all the possible samples that can be chosen.

c Find the sampling distribution for the mean $\overline{X} = \dfrac{X_1 + X_2}{2}$

Ⓟ **6** A manufacturer makes three sizes of toaster. 40% of the toasters sell for \$16, 50% sell for \$20, and 10% sell for \$30.

a Find the mean and variance of the value of the toasters.

A sample of two toasters is sent to a shop.

b List all the possible prices of the samples that could be sent.

c Find the sampling distribution for the mean price $\overline{X}$ of these samples.

Ⓟ **7** A supermarket sells a large number of 3-litre and 2-litre cartons of milk. They are sold in the ratio 3 : 2

a Find the mean and variance of the milk content in this population of cartons.

A random sample of 3 cartons is taken from the shelves (X_1, X_2 and X_3).

b List all the possible samples.

c Find the sampling distribution of the mean $\overline{X}$.

d Find the sampling distribution of the mode M.

e Find the sampling distribution of the median N of these samples.

Chapter review 6

1 A doctor's surgery is to offer health checks to all its patients over age 65. In order to estimate the amount of time needed to perform these health checks, the doctor decides to do the health check for a random sample of 20 patients over 65.

a Write down a suitable sampling frame that the doctor might use.

b Identify the sampling units.

2 The owners of a gym would like to change the gym's opening hours. They want to know if members would be happy with the new hours. They ask a random sample of 30 members.

a Write two likely reasons why the owners did not ask all the members.

b Suggest a suitable sampling frame.

c Identify the sampling units.

3 **a** Write down a reason why a sampling frame and a population may not be the same.

b Explain briefly why a sample is often used rather than a census.

4 **a** Explain what a statistic is.

A random sample $Y_1, Y_2, \ldots, Y_n$ is taken from a population with unknown mean μ.

b For each of the following, state with a reason whether or not it is a statistic.

i $\frac{Y_1 + Y_2 + Y_3}{4}$ **ii** $\frac{\sum Y}{n} - \mu$

5 A company manufactures electric light bulbs. They wish to see how many hours the light bulbs will work before failing. The company decides to test every 200th light bulb coming off the assembly line.

a Write down why the company does not test every light bulb.

b Identify the sampling units.

6 A call centre has 400 people (known as operators) operating the telephones. The manager decides that he needs to know how long the operators are spending on each call. He times a random sample of 30 operators over one day and works out the mean time per call.

a Write down two advantages of using a sample rather than a census in this case.

b Write down one disadvantage of using a sample in this case.

A sample is to be taken.

c Suggest a sampling frame.

d Identify the sampling units.

e Is the mean time the manager works out from the sample a statistic?

Give a reason for your answer.

(P) **7** A flower shop has ten florists (people who work in a flower shop). The owner wants to know if the florists are happy with the quality of the flowers being delivered to the shop. The owner asks all the florists for their views. Write down two reasons why the owner of the florist shop used a census.

(P) **8** The weights of tomatoes in a greenhouse are assumed to have mean μ and standard deviation σ.

A sample of 20 tomatoes were each weighed and their weights were recorded. If the sample is represented by $X_1, X_2, \ldots, X_{20}$ state whether or not the following are statistics.

a $\frac{X_1 + X_{20}}{3}$ **b** $\frac{\sum X}{20}$ **c** $\sum X^2 + \mu$ **d** $\frac{\sum X^2}{20} - \sigma^2$

(P) **9** A large box of coins contains 5p, 10p, and 20p coins in the ratio 3:2:1

a Find the mean μ and the variance σ^2 of the value of the coins.

A random sample of two coins is taken from the box and their values Y_1 and Y_2 are recorded.

b List all the possible samples that can be taken.

c Find the sampling distribution for the mean $(\bar{Y})$.

10 A bag contains a large number of counters:

60% have a value of 6

40% have a value of 10.

A random sample of three counters is drawn from the bag.

a Write down all the possible samples.

b Find the sampling distribution for the median N.

c Find the sampling distribution for the mode M.

Challenge

SKILLS CREATIVITY; ADAPTIVE LEARNING

Let $X_1, X_2, \ldots, X_n$ be n independent observations from $X \sim \mathrm{N}(\mu, \sigma^2)$

a Show that $\mathrm{E}(\overline{X}) = \mu$

b For $n = 3$, show that $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{3}$

Summary of key points

1 A population is a collection of individual items.

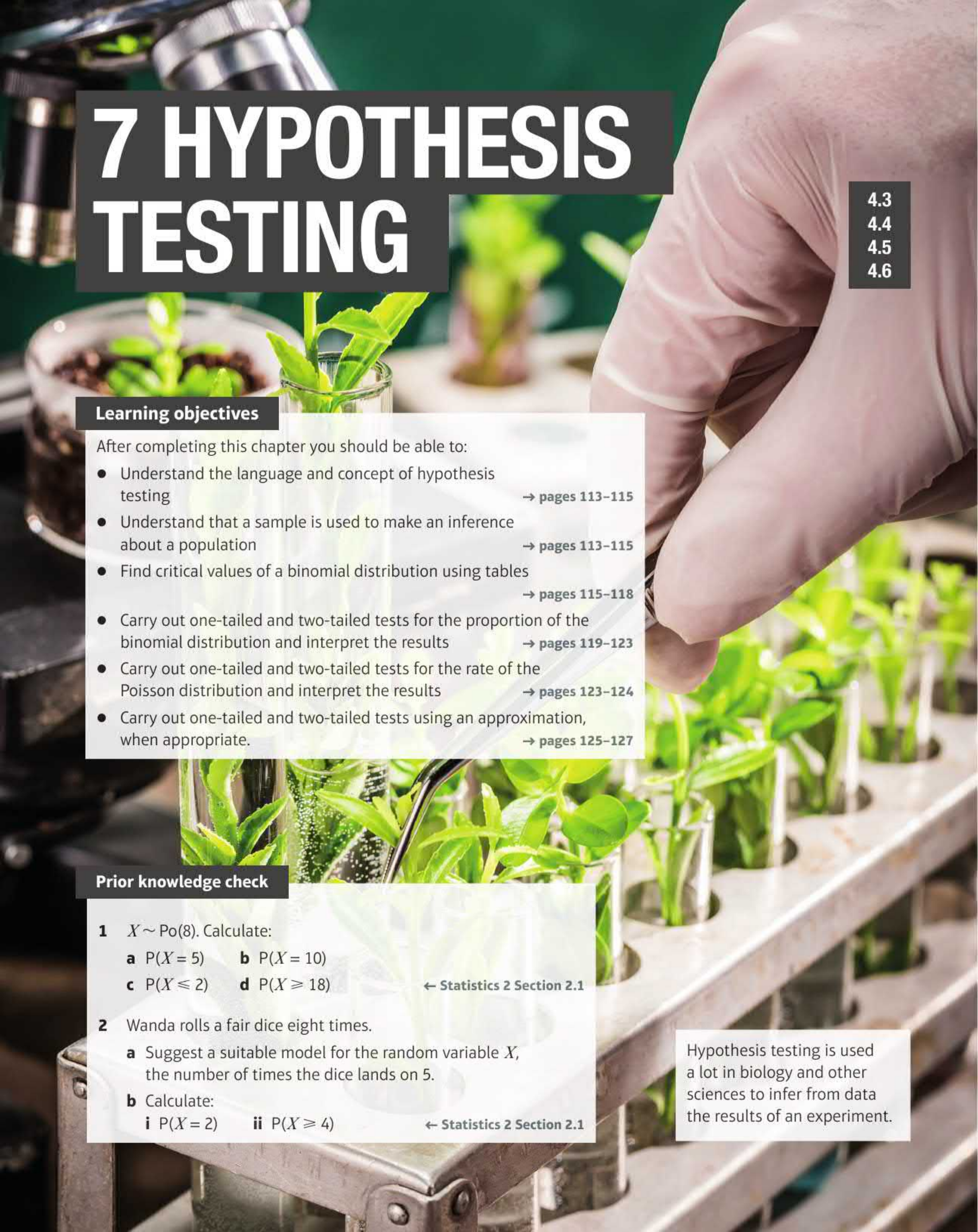
2 A sample is a selection of individual members or items from a population.

3 A sampling unit is an individual member of a population.

4 A sampling frame is a list of sampling units used in practice to represent a population.

5 A statistic is a quantity calculated only from the observations in a sample.

6 A statistic has a sampling distribution that is defined by giving all possible values of the statistic and the probability of each occurring.



7 HYPOTHESIS TESTING

4.3
4.4
4.5
4.6

Learning objectives

After completing this chapter you should be able to:

- Understand the language and concept of hypothesis testing → pages 113–115
- Understand that a sample is used to make an inference about a population → pages 113–115
- Find critical values of a binomial distribution using tables → pages 115–118
- Carry out one-tailed and two-tailed tests for the proportion of the binomial distribution and interpret the results → pages 119–123
- Carry out one-tailed and two-tailed tests for the rate of the Poisson distribution and interpret the results → pages 123–124
- Carry out one-tailed and two-tailed tests using an approximation, when appropriate. → pages 125–127

Prior knowledge check

1 $X \sim \text{Po}(8)$. Calculate:

a $\text{P}(X = 5)$ **b** $\text{P}(X = 10)$

c $\text{P}(X \leqslant 2)$ **d** $\text{P}(X \geqslant 18)$ ← Statistics 2 Section 2.1

2 Wanda rolls a fair dice eight times.

a Suggest a suitable model for the random variable X, the number of times the dice lands on 5.

b Calculate:
i $\text{P}(X = 2)$ **ii** $\text{P}(X \geqslant 4)$ ← Statistics 2 Section 2.1

Hypothesis testing is used a lot in biology and other sciences to infer from data the results of an experiment.

7.1 Hypothesis testing

A hypothesis is a statement made about the value of a **population parameter**. You can test a hypothesis about a population by carrying out an experiment or taking a sample from the population.

The result of the experiment or the statistic that is calculated from the sample is called the **test statistic**.

Links In this chapter, the population parameter you will be testing will be the probability p in a binomial distribution B(n, p), and the rate λ in a Poisson distribution $X \sim$ Po(λ).
← Statistics 2 Section 1.1

In order to carry out the test, you need to form two hypotheses:

- The **null hypothesis** H_0 is the hypothesis that you assume to be correct.
- The **alternative hypothesis** H_1 tells you about the parameter if your assumption is shown to be wrong.

Notation In this section, you should always give H_0 and H_1 in terms of the population parameter p when we are testing the probability using the binomial distribution, and λ when we are testing the rate using the Poisson distribution .

Example 1 SKILLS INTERPRETATION

Olav wants to see if a coin is unbiased or if it is biased in favour of landing on heads. He tosses the coin 8 times and counts the number of times X that it lands on heads.

a Describe the test statistic.

b Write down a suitable null hypothesis.

c Write down a suitable alternative hypothesis.

a The test statistic is X, the number of heads in 8 tosses.

The test statistic is calculated from the sample or experiment.

b If the coin is unbiased, the probability of it landing on heads is 0.5 so H_0: $p = 0.5$ is the null hypothesis.

You always write the null hypothesis in terms of the parameter. In this case, this is H_0: $p = \ldots$

c If the coin is biased in favour of landing on heads then the probability of landing on heads will be greater than 0.5 H_1: $p > 0.5$ is the alternative hypothesis.

If you were testing the coin for bias towards tails your alternative hypothesis would be H_1: $p < 0.5$

If you were testing the coin for bias in either direction your alternative hypothesis would be H_1: $p \neq 0.5$

- Hypothesis tests with alternative hypotheses in the form H_1: $p < \ldots$ and H_1: $p > \ldots$ are called one-tailed tests.
- Hypothesis tests with an alternative hypothesis in the form H_1: $p \neq \ldots$ are called two-tailed tests.

Hint You can think of a two-tailed test such as H_1: $p \neq 0.5$ as **two tests**, H_1: $p > 0.5$ or $p < 0.5$

To carry out a hypothesis test, you **assume the null hypothesis is true**. You then consider how likely it was for the observed value of the test statistic to occur. If this **likelihood** is less than a given threshold (called the **significance level** of the test) then you reject the null hypothesis.

Typically the significance level for a hypothesis test will be 10%, 5% or 1%, but you will be told which level to use in the question.

Example 2 SKILLS INTERPRETATION; REASONING

An election candidate believes she has the support of 40% of the residents in a particular town. A researcher wants to test, at the 5% significance level, whether the candidate is **over-estimating** her support. The researcher asks 20 people whether or not they support the candidate. Three people say that they support the candidate.

a Write down a suitable test statistic.

b Write down two suitable hypotheses.

c Explain the condition under which the null hypothesis would be rejected.

a The test statistic is the number of people who say they support the candidate.

b H_0: $p = 0.4$ H_1: $p < 0.4$

c The null hypothesis will be rejected if the probability of 3 or fewer people saying they support the candidate is less than 5%, given that $p = 0.4$

This is a one-tailed test; the researcher wants to see if the candidate is **over-estimating** her support. If she is, then the actual proportion of residents who support her will be **less than** 40%.

Watch out You are testing to see whether the actual probability is **less** than 0.4, so you would need to calculate the probability that the observed value of the test statistic is 3 or **fewer**.

Exercise 7A SKILLS INTERPRETATION

1 **a** Explain what you understand by a hypothesis test.

b Define a null hypothesis and an alternative hypothesis and state the symbols used for each.

c Define a test statistic.

2 For each of these hypotheses, state whether the hypotheses given describe a one-tailed or a two-tailed test:

a H_0: $p = 0.8$, H_1: $p > 0.8$ **b** H_0: $p = 0.6$, H_1: $p \neq 0.6$ **c** H_0: $p = 0.2$, H_1: $p < 0.2$

3 Dmitri wants to see if a dice is biased in favour of landing on 6. He throws the dice 60 times and counts the number of times that a 6 appears.

a Describe the test statistic.

b Write down a suitable null hypothesis to test this dice.

c Write down a suitable alternative hypothesis to test this dice.

Hint If the dice is biased in favour of landing on 6 then the probability of landing on 6 will be greater than $\frac{1}{6}$.

(P) **4** Isabelle wants to test to see if a coin is biased. She tosses the coin 100 times and counts the number of times she gets a head. Isabelle says that her test statistic is the probability of the coin landing on heads.

a Explain the mistake that Isabelle has made and state the correct test statistic for her test.

b Write down a suitable null hypothesis to test this coin.

c Write down a suitable alternative hypothesis to test this coin.

(P) **5** In a manufacturing process, the proportion p of faulty articles has been found, from long experience, to be 0.1. A sample of 100 articles from a new manufacturing process is tested, and 8 are found to be faulty. The manufacturers wish to test, at the 5% level of significance, whether or not there has been a reduction in the proportion of faulty articles.

a Suggest a suitable test statistic.

b Write down two suitable hypotheses.

c Explain the condition under which the null hypothesis is rejected.

(P) **6** In a TV talent show, a pop group gained 55% of the phone-in votes. In the next round, their singing was not so good. The TV company asked 20 people if they would still vote for the group, and 7 people said that they would. The TV company wants to test, at the 2% level of significance, if the level of support for the pop group has reduced.

a Write down a suitable test statistic.

b Write down two suitable hypotheses.

c Explain the condition under which the null hypothesis would be accepted.

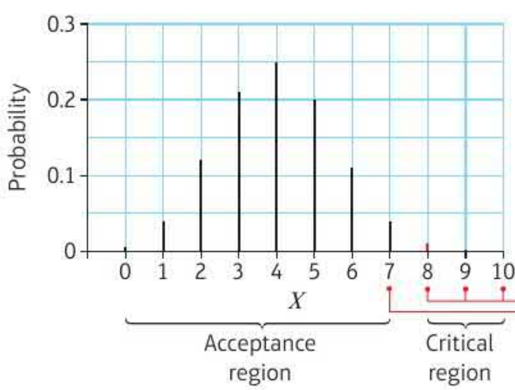
7.2 Finding critical values

When you carry out a hypothesis test, you need to be able to calculate the probability of the test statistic taking particular values, given that the null hypothesis is true.

In this chapter you will assume that the test statistic can be modelled by a binomial distribution. You will use this to calculate probabilities and find **critical regions**.

- **A critical region is a region of the probability distribution which, if the test statistic falls within it, would cause you to reject the null hypothesis.**

A test statistic is modelled as B(10, p) and a hypothesis test at the 5% significance level uses H_0: $p = 0.4$, H_1: $p > 0.4$. Assuming H_0 to be true, X has the following distribution: $X \sim$ B(10, 0.4)



P($X = 10$) = 0.0001, P($X = 9$) = 0.0016 and P($X = 8$) = 0.0106. Hence P($X \geqslant 8$) = 0.0123. This is less than the significance level of 5%. A test statistic of 10, 9 or 8 would lead to the null hypothesis being rejected.

P($X = 7$) = 0.0425. Adding this probability to P($X \geqslant 8$) takes the probability over 0.05, so a test statistic of 7 or fewer would lead to the null hypothesis being accepted.

- The **critical value** is the first value to fall inside of the critical region.

In this example, the critical value is 8 and the critical region is 8, 9 or 10.
7 falls in the **acceptance region**, the region where we accept the null hypothesis.

The critical value and hence the critical region can be determined from binomial distribution tables, or by finding cumulative binomial probabilities using your calculator.

In the $n = 10$ table, the critical region is found by looking for a probability such that $P(X \geqslant x) < 0.05$

$p =$	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 10, x = 0$	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
1	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
2	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
3	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
4	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
5	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
6	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
7	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
8	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990

$P(X \leqslant 6) = 0.9452$ so
$P(X \geqslant 7) = 0.0548$ (>0.05)

$P(X \leqslant 7) = 0.9877$ so
$P(X \geqslant 8) = 0.0132$ (<0.05)

This shows that $x = 7$ is not extreme enough to lead to the rejection of the null hypothesis but that $x = 8$ is. Hence $x = 8$ is the critical value and $x \geqslant 8$ is the critical region.

The probability of the test statistic falling within the critical region, given that H_0 is true, is 0.0132 or 1.32%. This is sometimes called the **actual significance level** of the test.

- The actual significance level of a hypothesis test is the probability of incorrectly rejecting the null hypothesis.

Watch out The threshold probability for your test (1%, 5%, 10%) is often referred to as the level of significance for your test. This might be different from the actual significance level, which is the probability that your test statistic would fall within the critical region even if H_0 is true.

Example 3 SKILLS ANALYSIS

A single observation is taken from a binomial distribution $B(6, p)$.
The observation is used to test $H_0: p = 0.35$ against $H_1: p > 0.35$

a Using a 5% level of significance, find the critical region for this test.

b State the actual significance level of this test.

a Assume H_0 is true. Then $X \sim B(6, 0.35)$

$P(X \geqslant 4) = 1 - P(X \leqslant 3) = 1 - 0.8826 = 0.1174$

$P(X \geqslant 5) = 1 - P(X \leqslant 4) = 1 - 0.9777 = 0.0223$

The critical region is 5 or 6.

Use tables or your calculator to find the first value of x for which $P(X \geqslant x) < 0.05$

$P(X \geqslant 4) > 0.05$ but $P(X \geqslant 5) < 0.05$, so 5 is the critical value.

b The actual significance level is the probability of incorrectly rejecting the null hypothesis:
$P(\text{reject null hypothesis}) = P(X \geqslant 5)$
$= 0.0223$
$= 2.23\%$

This is the same as the probability that X falls within the critical region.

- **For a two-tailed test there are two critical regions: one at each end of the distribution.**

Example 4

SKILLS ADAPTIVE LEARNING

A random variable X has binomial distribution $B(40, p)$.
A single observation is used to test H_0: $p = 0.25$ against H_1: $p \neq 0.25$

a Using the 2% level of significance, find the critical region of this test. The probability in each tail should be as close as possible to 0.01

b Write down the actual significance level of the test.

a Assume H_0 is true. Then $X \sim B(40, 0.25)$
Consider the lower tail:
$P(X \leqslant 4) = 0.0160$
$P(X \leqslant 3) = 0.0047$
Consider the upper tail:
$P(X \geqslant 19) = 1 - P(X \leqslant 18) = 1 - 0.9983$
$= 0.0017$
$P(X \geqslant 18) = 1 - P(X \leqslant 17) = 1 - 0.9884$
$= 0.0116$
The critical regions are $0 \leqslant X \leqslant 3$ and $18 \leqslant X \leqslant 40$.

b The actual significance level is
$0.0047 + 0.0116 = 0.0163 = 1.63\%$

$P(X \leqslant 3)$ is closest to 0.01, so 3 is the critical value for this tail.

Watch out Read the question carefully. Even though $P(X \geqslant 18)$ is greater than 0.01 it is still the closest value to 0.01. The critical value for this tail is 18.

Online Use technology to explore the locations of the critical values for each tail in this example.

Exercise 7B

SKILLS ANALYSIS

1 Explain what you understand by the following terms:
a critical value **b** critical region **c** acceptance region.

2 A test statistic has a distribution $B(10, p)$. Given that H_0: $p = 0.2$, H_1: $p > 0.2$, find the critical region for the test using a 5% significance level.

3 A random variable has a distribution $B(20, p)$. A single observation is used to test H_0: $p = 0.15$ against H_1: $p < 0.15$. Using a 5% level of significance, find the critical region of this test.

(E) **4** A random variable has distribution $B(20, p)$. A single observation is used to test H_0: $p = 0.4$ against H_1: $p \neq 0.4$

a Using a 5% level of significance, find the critical region of this test. **(3 marks)**

b Write down the actual significance level of the test. **(1 mark)**

5 A test statistic has a distribution B(20, p). Given that H_0: $p = 0.15$, H_1: $p < 0.15$, find the critical region for the test using a 1% level of significance.

E/P 6 A random variable has distribution B(10, p). A single observation is used to test H_0: $p = 0.2$ against H_1: $p \neq 0.2$

a Using a 1% level of significance, find the critical region of this test. The probability in each tail should be as close as possible to 0.005 **(3 marks)**

b Write down the actual significance level of the test. **(2 marks)**

E/P 7 A mechanical component fails, on average, 3 times out of every 10. An engineer suggests a new method of making the component that he believes reduces the likelihood of failure. He tests a sample of 20 components made using his new method.

a Describe the test statistic. **(1 mark)**

b State suitable null and alternative hypotheses. **(2 marks)**

c Using a 5% level of significance, find the critical region for a test to check his belief, ensuring the probability is as close as possible to 0.05 **(3 marks)**

d Write down the actual significance level of the test. **(1 mark)**

E/P 8 Seedlings come in trays of 40. On average, 12 seedlings survive to be planted on. A gardener decides to use a new fertiliser on the seedlings which she believes will improve the number that survive.

a Describe the test statistic and state suitable null and alternative hypotheses. **(3 marks)**

b Using a 10% level of significance, find the critical region for a test to check her belief. **(3 marks)**

c State the probability of incorrectly rejecting H_0 using this critical region. **(1 mark)**

E/P 9 A restaurant owner notices that her customers typically choose pasta one fifth of the time. She changes the recipe and believes this will change the proportion of customers choosing pasta.

a Suggest a model and state suitable null and alternative hypotheses. **(3 marks)**

The restaurant owner takes a random sample of 25 customers.

b Find, at the 5% level of significance, the critical region for a test to check her belief. **(4 marks)**

c State the probability of incorrectly rejecting H_0. **(1 mark)**

Challenge

A test statistic has binomial distribution B(50, p).

Given that H_0: $p = 0.7$, H_1: $p \neq 0.7$:

a find the critical region for the test statistic such that the probability in each tail is close as possible to 5%.

Chloe takes two observations of the test statistic and finds that they both fall inside the critical region. Chloe decides to reject H_0.

b Find the probability that Chloe has incorrectly rejected H_0.

7.3 One-tailed tests

If you have to carry out a one-tailed hypothesis test, you need to:

- Formulate a model for the test statistic
- Identify suitable null and alternative hypotheses
- Calculate the probability of the test statistic taking the observed value (or higher/lower), assuming the null hypothesis is true
- Compare this to the significance level
- Write a conclusion in the context of the question

Alternatively, you can find the critical region and see if the observed value of the test statistic lies inside it.

Example 5

The standard treatment for a particular disease has a $\frac{2}{5}$ probability of success. A certain doctor has undertaken research in this area and has produced a new medicine which has been successful with 11 out of 20 patients. The doctor claims that the new medicine represents an improvement on the standard treatment.

Test, at the 5% significance level, the claim made by the doctor.

X represents the number of patients in the trial for whom the drug was successful.

p is the probability of success for each patient.

Define your test statistic X and parameter p.

$X \sim B(20, p)$

$H_0: p = 0.4 \qquad H_1: p > 0.4$

Write down the model for your test statistic and your hypotheses. The doctor claims the drug represents an improvement so the alternative hypothesis is $p > 0.4$

Method 1

Assume H_0 is true. Then $X \sim B(20, 0.4)$

Assume the null hypothesis is true and calculate the probability of 11 or more successful treatments.

$P(X \geqslant 11) = 1 - P(X \leqslant 10)$

$= 1 - 0.8725$

Use the cumulative binomial tables or your calculator to find $P(X \leqslant 10)$.

$= 0.1275$

$= 12.75\%$

12.75% > 5% so there is not enough evidence to reject H_0.

Compare the probability to the significance level of your test.

The new drug is no better than the old one.

Make sure you write a conclusion in context.

Method 2

$P(X \geqslant 13) = 1 - P(X \leqslant 12) = 0.021$

$P(X \geqslant 12) = 1 - P(X \leqslant 11) = 0.0565$

Work out the critical region and see if 11 lies within it.

The critical region is 13 or more.

Since 11 does not lie in the critical region, we accept H_0.

There is insufficient evidence that the new drug is better than the old one.

Unless you are specifically instructed as to which method to use, you can use the one you prefer.

Exercise 7C

SKILLS ANALYSIS; DECISION MAKING

1 A single observation x is taken from a binomial distribution $B(10, p)$ and a value of 5 is obtained. Use this observation to test $H_0: p = 0.25$ against $H_1: p > 0.25$ using a 5% significance level.

2 A random variable has distribution $X \sim B(10, p)$. A single observation of $x = 1$ is taken from this distribution. Test, at the 5% significance level, $H_0: p = 0.4$ against $H_1: p < 0.4$

3 A single observation x is taken from a binomial distribution $B(20, p)$ and a value of 10 is obtained. Use this observation to test $H_0: p = 0.3$ against $H_1: p > 0.3$ using a 5% significance level.

4 A random variable has distribution $X \sim B(20, p)$. A single observation of $x = 3$ is taken from this distribution. Test, at the 1% significance level, $H_0: p = 0.45$ against $H_1: p < 0.45$

5 A single observation x is taken from a binomial distribution $B(20, p)$ and a value of 2 is obtained. Use this observation to test $H_0: p = 0.25$ against $H_1: p < 0.25$ using a 5% significance level.

6 A random variable has distribution $X \sim B(8, p)$. A single observation of $x = 7$ is taken from this distribution. Test, at the 5% significance level, $H_0: p = 0.3$ against $H_1: p > 0.3$

(P) 7 While playing a board game, some players think the dice isn't landing on the number 6 often enough. During a particular game, the dice was rolled 12 times and only one 6 appeared. Does this represent sufficient evidence, at the 5% level of significance, that the probability of landing on a 6 is less than $\frac{1}{6}$?

(P) 8 The success rate of the standard treatment for patients suffering from a particular skin disease is claimed to be 65%.

a In a sample of n patients, X is the number for which the treatment is successful. Write down a suitable distribution to model X. Give reasons for your choice of model.

A random sample of 10 patients receives the standard treatment and in only 3 cases was the treatment successful. It is thought that the standard treatment was not as effective as it is claimed.

b Test the claim at the 5% level of significance.

(E/P) 9 A seed planting method is successful on average 4 times out of every 10. A horticulturist develops a new technique which she believes will improve the number of plants that successfully grow. She takes a random sample of 20 seeds and attempts to grow them.

a Using a 5% level of significance, find the critical region for a test to check her belief. **(4 marks)**

b Of her sample of 20 plants, the horticulturist finds that 14 have grown. Comment on this observation in light of the critical region. **(2 marks)**

Problem-solving

In this question you are told to find the critical region in part **a**. You will save time by using your critical region to answer part **b**.

E/P **10** It is claimed that an election candidate has 35% support. It is revealed that the candidate aims to support local charities if elected. It is thought that the level of support will go up as a result. In a poll, 50 new voters are asked who they will vote for.

a Describe the test statistic and state suitable null and alternative hypotheses. **(2 marks)**

b Using a 5% level of significance, find the critical region for a test to check the belief. **(4 marks)**

c In the new poll, 28 people are found to support the candidate. Comment on this observation in light of the critical region. **(2 marks)**

7.4 Two-tailed tests

In section 7.3, all of the hypothesis tests were testing whether the proportion had gone up or had gone down. There is a direction to H_1.

Sometimes we may be asked if it has changed. We are not told how it has changed and so there is no direction to H_1. This is called a two-tailed test.

- For a two-tailed test, halve the significance level at the end you are testing.

You need to know which tail of the distribution you are testing. If the test statistic is $X \sim B(n, p)$ then the **expected** outcome is np. If the observed value x is lower than this then consider $P(X \leqslant x)$. If the observed value is higher than the expected value, then consider $P(X \geqslant x)$. In your exam it will usually be obvious which tail you should test.

Example 6 SKILLS DECISION MAKING

Over a long period of time, it has been found that at Enrico's restaurant the ratio of non-vegetarian to vegetarian meals is 2 to 1. At Manuel's restaurant, in a random sample of 10 people ordering meals, only one ordered a vegetarian meal. Using a 5% level of significance, test whether or not the proportion of people eating vegetarian meals at Manuel's restaurant is different to that at Enrico's restaurant.

The proportion of people eating vegetarian meals at Enrico's is $\frac{1}{3}$.

X is the number of people in the sample at Manuel's who order vegetarian meals.

p is the probability that a randomly chosen person at Manuel's orders a vegetarian meal.

$H_0 : p = \frac{1}{3}$ $H_1 : p \neq \frac{1}{3}$

Significance level 5%

If H_0 is true, then $X \sim B(10, \frac{1}{3})$

Problem-solving

You can use the same techniques to hypothesis-test for the **proportion** of a population that have a given characteristic. You could equivalently define the test parameter as the proportion of diners at Manuel's that order a vegetarian meal.

Hypotheses. The test will be two-tailed as we are testing if they are different.

Method 1

$P(X \leqslant 1) = P(X = 0) + P(X = 1)$

$= \left(\frac{2}{3}\right)^{10} + 10\left(\frac{2}{3}\right)^{9}\left(\frac{1}{3}\right)$

$= 0.017\,34... + 0.086\,70...$

$= 0.104$ (3 s.f.)

$0.104 > 0.025$

There is insufficient evidence to reject H_0.

There is no evidence that the proportion of vegetarian meals at Manuel's restaurant is different to Enrico's.

Method 2

Let c_1 and c_2 be the two critical values.

$P(X \leqslant c_1) \leqslant 0.025$ and $P(X \geqslant c_2) \leqslant 0.025$

For the lower tail:

$P(X \leqslant 0) = 0.017\,341... < 0.025$

$P(X \leqslant 1) = 0.104\,04... > 0.025$

So $c_1 = 0$

For the upper tail:

$P(X \geqslant 6) = 1 - P(X \leqslant 5)$

$= 0.076\,56... > 0.025$

$P(X \geqslant 7) = 1 - P(X \leqslant 6)$

$= 0.019\,66... < 0.025$

So $c_2 = 7$

The observed value of 1 does not lie in the critical region, so H_0 is not rejected. There is no evidence that the proportion of people eating vegetarian meals has changed.

> The expected value would be $10 \times \frac{1}{3} = 3.333...$ The observed value, 1, is less than this so consider $P(X \leqslant 1)$.

> You can calculate the probabilities long-hand, like this, or use your calculator with $p = 0.3333$

> We use 0.025 because the test is two-tailed.

> Conclusion and what it means in context.

> The probability in each critical region should be less than $0.05 \div 2 = 0.025$

> Use the cumulative binomial function on your calculator, with $n = 10$ and $p = 0.3333$

> Write down a value on either side of the boundary to show that you have determined the correct critical values.

> Remember to write a conclusion in the context of the question.

Exercise 7D

SKILLS DECISION MAKING

1 A single observation x is taken from a binomial distribution B(30, p) and a value of 10 is obtained. Use this observation to test H_0: $p = 0.5$ against H_1: $p \neq 0.5$ using a 5% significance level.

2 A random variable has distribution $X \sim B(25, p)$. A single observation of $x = 10$ is taken from this distribution. Test, at the 10% significance level, H_0: $p = 0.3$ against H_1: $p \neq 0.3$

3 A single observation x is taken from a binomial distribution B(10, p) and a value of 9 is obtained. Use this observation to test H_0: $p = 0.75$ against H_1: $p \neq 0.75$ using a 5% significance level.

4 A random variable has distribution $X \sim B(20, p)$. A single observation of $x = 1$ is taken from this distribution. Test, at the 1% significance level, H_0: $p = 0.6$ against H_1: $p \neq 0.6$

(P) **5** A random variable has distribution $X \sim B(50, p)$. A single observation of $x = 4$ is taken from this distribution. Test, at the 2% significance level, $H_0: p = 0.02$ against $H_1: p \neq 0.02$

Watch out Although the observed value of 4 appears to be small, the expected value of X is actually $50 \times 0.02 = 1$. You need to consider the upper tail of the distribution: $P(X \geqslant 4)$.

(P) **6** A coin is tossed 20 times, and lands on heads 6 times. Use a two-tailed test with a 5% significance level to determine whether there is sufficient evidence to conclude that the coin is biased.

(E/P) **7** The national proportion of people who get an infection after having a particular operation in hospitals is 20%. A hospital decides to take a sample of size 20 from their records.

a Find critical regions, at the 5% level of significance, to test whether or not their proportion of infections differs from the national proportion. The probability in each tail should be as close to 2.5% as possible. **(5 marks)**

b State the actual significance level of the test. **(1 mark)**

The hospital finds that 8 of their 20 patients experienced infection.

c Comment on this finding in light of your critical regions. **(2 marks)**

(E/P) **8** A machine makes glass bowls. It is observed that one in ten of the bowls have small cracks in them. The production process is improved and a sample of 20 bowls is taken. One of the bowls is cracked. Test, at the 10% level of significance, the hypothesis that the proportion of cracked bowls has changed as a result of the change in the production process. State your hypotheses clearly. **(7 marks)**

(E/P) **9** Over a period of time, Agnetha has discovered that the carrots she grows have a 25% chance of being longer than 7 cm. She tries a new type of fertiliser to help them grow. In a random sample of 30 carrots, 13 are longer than 7 cm. Agnetha claims that the new fertiliser has changed the probability of a carrot being longer than 7 cm. Test Agnetha's claim at the 5% significance level. State your hypotheses clearly. **(7 marks)**

(E/P) **10** A standard blood test is able to identify a particular disease with probability 0.96. A manufacturer suggests that a cheaper test will have the same probability of success. It conducts a trial on 75 patients. The new test correctly identifies the disease in 63 of these patients. Test the manufacturer's claim at the 10% level, stating your hypotheses clearly. **(7 marks)**

7.5 Testing the mean λ of a Poisson distribution

In the previous section, we saw how the binomial distribution is used to test for a probability or a proportion.

When conducting a hypothesis test, it is important to know the underlying distribution and the population parameter that is to be tested.

We can test the mean λ of a Poisson distribution in a similar way to the binomial since it is also a discrete distribution.

Example 7 SKILLS ADAPTIVE LEARNING

An office manager finds that, over a long time, incoming telephone calls from customers occur at a rate of 0.325 per minute.

The manager believes that the number of calls has increased recently. To test this, the number of incoming calls during a random 20-minute interval is recorded.

a Find the critical region for a two-tailed test of the hypothesis that the number of incoming calls occur at the rate of 0.325 per minute. The probability in both tails should be as close to 2.5% as possible.

b Write down the actual significance level of the above test.

Later, the office manager runs an advertising campaign and records 1 call in a 10-minute interval.

c Test, at the 5% level of significance, whether or not there is evidence that the rate of incoming calls has decreased.

a Let X represent the number of telephone calls in a 20-minute interval. Then $X \sim \text{Po}(\lambda)$

The rate is $20 \times 0.325 = 6.5$ and so

$H_0: \lambda = 6.5$

$H_1: \lambda \neq 6.5$

> The hypothesis can be stated in terms of the rate. In terms of the rate, the hypotheses would be:
> $H_0: \lambda = 0.325$
> $H_1: \lambda \neq 0.325$

Consider the lower tail:

$P(X \leqslant 2) = 0.0430$

$P(X \leqslant 1) = 0.0113$

> $P(X \leqslant 1)$ is closest to 0.025, so 1 is the critical value for this tail.

Consider the upper tail:

$P(X \geqslant 12) = 1 - P(X \leqslant 11) = 0.0339$

$P(X \geqslant 13) = 1 - P(X \leqslant 12) = 0.0160$

> $P(X \geqslant 12)$ is closest to 0.025, so 12 is the critical value for this tail.

The critical regions are $0 \leqslant X \leqslant 1$ and $X \geqslant 12$

b The actual significance level is:

$0.0113 + 00339 = 0.0452$

c Let Y represent the number of telephone calls in a 10-minute interval. Then $Y \sim \text{Po}(\lambda)$

The rate is $10 \times 0.325 = 3.25$ and so

$H_0: \lambda = 3.25$

$H_1: \lambda < 3.25$

> We are not able to use the cumulative tables here since 3.25 is not in the cumulative tables.

This is a one-tailed test.

$P(Y \leqslant 1) = P(Y = 0) + P(Y = 1)$

$= e^{-3.25} + e^{-3.25} \times 3.25$

$= 0.1648$

> Calculate $P(Y \leqslant \text{test statistic})$

Since $0.1648 < 0.05$, the test statistic is not in the critical region.

> $(Y \leqslant \text{test statistic}) < \text{significance level}$, so we do not reject the null hypothesis.

Do not reject H_0

There is insufficient evidence to suggest that the rate of telephone calls has decreased.

7.6 Using approximations

Whether you are using critical regions or not, with both the binomial and Poisson distributions, the number in the sample could be large in practice. This makes the calculations very difficult. It is possible in these cases, if the conditions are suitable, to use an approximation, as shown in the following examples.

Example 8

SKILLS ADAPTIVE LEARNING; DECISION MAKING

A shop sells cars at a rate of 10 per week. In an attempt to increase sales, car prices were reduced for a six week period. During this period a total of 75 cars were sold.

Using a 5% level of significance, test whether or not there is evidence that the average number of sales per week has increased during this six-week period.

$H_0: \lambda = 10$

$H_1: \lambda > 10$

We use the parameter as $\lambda = 10$ since we are testing the **rate per week**.

Let Y represent the number of sales in a six-week period. Under the hypothesis $H_0: \lambda = 10$

$Y \sim \text{Po}(60)$

$P(Y \geqslant 75) = 1 - P(Y \leqslant 74)$

$\approx 1 - P(W \leqslant 74.5)$ where $W \sim N(60, 60)$

Don't forget the continuity correction.

$\approx 1 - P\left(Z < \frac{74.5 - 60}{\sqrt{60}}\right) = 0.0307$

$0.0307 < 0.05$, therefore reject H_0

There is sufficient evidence to suggest that the number of sales per week has increased.

Remember to write the conclusion in context.

Example 9

A manager thinks that her sales staff make a sale to 45% of customers entering their shop. She randomly selects 100 customers. Of the 100 customers, 35 were sold something.

By using a suitable approximation, test, at the 5% level of significance, whether or not the manager's claim is justified.

$H_0: p = 0.45$

$H_1: p \neq 0.45$

Let X represent the number of customers who were sold something, and under H_0

$X \sim B(100, 0.45)$

$P(X \leqslant 35) \approx P(W < 35.5)$ where $W \sim N(45, 24.75)$

$$\approx P\left(Z < \frac{35.5 - 45}{\sqrt{24.75}}\right) = 0.0281$$

Since $0.0281 > 0.025$ we do not reject H_0

There is insufficient evidence that the manager is wrong.

Since n is large and $p \approx 0.5$ we can approximate this as a normal distribution, $W \sim N(np, np(1 - p))$

Example 10 SKILLS DECISION MAKING

During an outbreak of flu, 4% of the population of a large city was affected on a given day. The manager of a factory that employs 100 people found that 12 employees were absent, claiming they had caught the flu.

a Using a 5% level of significance, find the critical region that the manager could use to test whether or not more than 4% of her workers claimed to have caught the flu.

b State the conclusion the manager came to, giving a reason for your answer.

a H_0: $p = 0.04$

H_1: $p > 0.04$

We are testing a proportion and so the underlying distribution is binomial.

X represents the number of employees absent.

Under H_0:, $X \sim B(100, 0.04)$

$X \approx Po(4)$

Since n is large, and p is small, we can approximate this as a Poisson distribution.

$P(X \leqslant 9) = 1 - P(X \leqslant 8) = 1 - 0.9786 = 0.0214$

$P(X \geqslant 8) = 1 - P(X \leqslant 7) = 1 - 0.9489 = 0.0511$

The critical region is therefore $X \geqslant 9$

since $P(X \geqslant 9) = 0.0214 < 0.05$ we reject H_0

b Since $12 > 9$, the manager concluded that the percentage of employees claiming to have flu was greater than 4%.

Exercise 7E SKILLS ANALYSIS; DECISION MAKING

For questions 1 to 3, find the critical region for the test statistic X, given that X has a Po(4) distribution.

1 H_0: $\lambda = 4$ and H_1: $\lambda > 4$, using a 5% level of significance

2 H_0: $\lambda = 9$ and H_1: $\lambda < 9$, using a 5% level of significance

3 H_0: $\lambda = 3.5$ and H_1: $\lambda < 3.5$, using a 5% level of significance

4 A single observation x is taken from a Poisson distribution $X \sim Po(\lambda)$ and a value of 7 is obtained. Use this observation to test H_1: $\lambda = 5$ against H_1: $\lambda \neq 5$ using a 5% level of significance.

5 A random variable has distribution $X \sim \text{Po}(\lambda)$. A single observation of $x = 11$ is taken from this distribution. Test, at the 10% significance level, $H_0: \lambda = 8$ against $H_1: \lambda \neq 8$

6 A single observation x is taken from a Poisson distribution $\text{Po}(\lambda)$ and a value of 4 is obtained. Use this observation to test $H_0: \lambda = 6$ against $H_1: \lambda < 6$, using a 5% level of significance.

7 A random variable has distribution $X \sim \text{Po}(\lambda)$. A single observation of $x = 8$ is taken from this distribution. Test, at the 10% significance level, $H_0: \lambda = 4$ against $H_1: \lambda > 4$

8 Over a number of years, the mean number of storms in a certain area during August is 4. A scientist suggests that, due to global warming, the number of storms will have increased. The scientist carries out a hypothesis test based on the number of storms.

a Suggest a suitable hypothesis for this test.

b Find to what level the number of storms must increase for the null hypothesis to be rejected at the 5% level of significance.

c The actual number of storms this August was 8. What conclusion did the scientist come to?

9 An estate agent usually sells apartments at a rate of 10 per week. Over an eight-week period during a recession (when money was less available) he sold 55 apartments.

Using a suitable approximation, test, at the 5% level of significance, whether or not there is evidence that the weekly rate of sales was reduced by the recession.

10 A manager thinks that 20% of his workforce are absent for at least one day each month. He chooses 100 workers at random and finds that in the last month, two employees had been absent for at least one day.

Using a suitable approximation, test, at the 5% level of significance, whether or not this provides evidence that the percentage of workers who are absent for at least 1 day per month is less than 20%.

Chapter review 7

E/P **1** Mai travels to work five days a week on a train. She does two journeys per day. Over a long period of time, she finds that the train is late 20% of the time. A new company takes over the train service that Mai uses. Mai thinks that the service will be late more often. In the first week of the new service the train is late three times. You may assume that the number of times the train is late in a week has a binomial distribution. Test, at the 5% level of significance, whether or not there is evidence that there is an increase in the number of times the train is late. State your hypothesis clearly. **(7 marks)**

E/P **2** A marketing company claims that Sunshine Tea tastes better than Blossom's Tea. Five people chosen at random as they entered a supermarket were asked to say which brand they preferred. Four people preferred Sunshine Tea. Test, at the 5% level of significance, whether or not the marketing company's claim is true. State your hypothesis clearly. **(7 marks)**

E/P **3** Records suggest that, nationally, 30% of cars fail a brake test.

a Give a reason to support the use of a binomial distribution as a suitable model for the number of cars failing a brake test. **(1 mark)**

b If 5 cars take the brake test, find the probability that all of them pass. **(2 marks)**

A garage decides to conduct a survey of their cars. A randomly selected sample of 10 of their cars is tested. Two of them fail the test.

c Test, at the 5% level of significance, whether or not there is evidence to support the suggestion that cars in this garage fail less than the national average. **(7 marks)**

E/P **4** The proportion of faulty articles in a certain manufacturing process has been found from long experience to be 0.1. A random sample of 50 articles was taken in order to monitor the production. The number of faulty articles was recorded.

a Using a 5% level of significance, find the critical regions for a two-tailed test of the hypothesis that 1 in 10 articles has a fault. The probability in each tail should be as close to 2.5% as possible. **(4 marks)**

b State the actual significance level of the above test. **(2 marks)**

Another sample of 20 articles was taken at a later date. Four articles were found to be faulty.

c Test, at the 10% significance level, whether or not there is evidence that the proportion of faulty articles has increased. State your hypothesis clearly. **(5 marks)**

E/P **5** It is claimed that 50% of people use Oriels washing powder. In a random survey of 20 people, 12 said they did not use Oriels washing powder.
Test, at the 5% significance level, whether or not there is evidence that the proportion of people using Oriels washing powder is 0.5. State your hypothesis carefully. **(6 marks)**

E/P **6** The manager of a superstore thinks that the probability of a person buying a certain model of computer is only 0.2
To test whether this hypothesis is true, the manager decides to record the model of computer bought by a random sample of 50 people who bought a computer.

a Find the critical region that would enable the manager to test whether or not there is evidence that the probability is different from 0.2.

The probability of each tail should be as close to 2.5% as possible. **(4 marks)**

b Write down the significance level of this test. **(2 marks)**

15 people buy that certain model.

c Comment on this observation in light of your critical region. **(2 marks)**

E/P **7** **a** Explain what is meant by:

i a hypothesis test **ii** a critical value **iii** an acceptance region. **(3 marks)**

Johan believes the probability of him being late to school is 0.2. To test this claim, he counts the number of times he is late in a random sample of 20 days.

b Find the critical region for a two-tailed test, at the 10% level of significance, of whether the probability he is late for school differs from 0.2 **(5 marks)**

c State the actual significance level of the test. **(1 mark)**

Johan discovers he is late for school in 7 out of the 20 days.

d Comment on whether Johan should accept or reject his claim that the probability he is late for school is 0.2 **(2 marks)**

E/P **8** From a large data set, the likelihood of a day with either zero or trace amounts of rain in Hamburg in June 1987 was 0.5

Clara believes that the likelihood of a rain-free day in 2015 has increased.

In June 2015 in Hamburg, 21 days were observed as having zero or trace amounts of rain.

Using a 5% significance level, test whether or not there is evidence to support Clara's claim. **(6 marks)**

E/P **9** A single observation x is to be taken from a binomial distribution B(30, p). This observation is used to test H_0: $p = 0.35$ against H_1: $p \neq 0.35$

a Using a 5% level of significance, find the critical region for this test. The probability of rejecting either tail should be as close as possible to 2.5% **(3 marks)**

b State the actual significance level of this test. **(2 marks)**

The actual value of X obtained is 4.

c State a conclusion that can be drawn based on this value, giving a reason for your answer. **(2 marks)**

E/P **10** A company that makes medicines claims that 85% of patients suffering from a severe rash (red spots on the skin) recover when treated with a new skin cream.

A random sample of 20 patients with this rash is taken from hospital records.

a Write down a suitable distribution to model the number of patients in this sample who recover when treated with the new cream. **(2 marks)**

b Given that the claim is correct, find the probability that the cream will be successful for exactly 16 patients. **(2 marks)**

The hospital believes that the claim is incorrect and the percentage who will recover is lower. From the records, a nurse took a random sample of 30 patients who had been prescribed the cream. She found that 20 had recovered.

c Stating your hypotheses clearly, test, at the 5% level of significance, the hospital's belief. **(6 marks)**

11 **a** Explain what you understand by a hypothesis test.

During a carnival, cups of coffee are thought to be sold at a rate of 2 every minute. To test this, the number of cups of coffee sold during a random 30-minute interval is recorded.

b State one reason why the sale of cups of coffee can be modelled by a Poisson distribution.

c Find the critical region for a two-tailed hypothesis that the number of cups of coffee sold occurs at a rate of 2 every minute. The probability in each tail should be as close to 2.5% as possible.

d Write down the actual significance level of the above test.

12 A large caravan company hires caravans out for a week at a time. During the winter, the mean number of caravans hired is 6 per week.

a Calculate the probability that in one particular week in winter, the company will hire out exactly 4 caravans.

The company decides to reduce prices in winter and do extra advertising. This results in the mean number of caravans being hired out rising to 11 per week.

b Test, at the 5% significance level, whether or not the proportion of caravans hired out has increased. State your hypothesis clearly.

13 At one stage of a water treatment process, the number of particles of dirt per litre present in the water has a Poisson distribution with mean 10. The water then enters a filter (something you pass a liquid through to remove particular substances) which should extract 75% of dirt.

The manager of the water treatment facility orders a study into the effectiveness of this filter. Twenty 1-litre samples are taken from the water and 64 particles of foreign matter are found.

Using a suitable approximation, test, at the 5% level of significance, whether or not there is evidence that the filter is failing to work properly.

14 A shop finds that it sells jars of onion marmalade at the rate of 10 per week.
During a television cookery programme, onion marmalade is used in a recipe.
Over the next six weeks the shop sells 84 jars of onion marmalade.

Using a suitable approximation, test, at the 5% level of significance, whether or not there is evidence that the rate of sales after the television programme has increased as a result of the recipe in the cookery programme.

15 A manufacturer produces large quantities of patterned plates. It is known from previous records that 6% of the plates will be rejected because of flaws in the pattern. To check that the production process is not getting worse, the manager takes a sample of 150 plates and finds that 15 have flaws in their pattern.

Use a suitable approximation to test, at the 5% significance level, whether or not the process is getting worse.

16 Julia grows apples. Over a period of time, she finds that the probability of an apple being below the size required by a supermarket is 0.45. She has recently planted another orchard using a different variety of apple. A sample of 200 of this other variety of apple had 60 rejected as being undersized.

Use a suitable approximation to test, at the 5% significance level, whether or not the new variety of apple is better than the old variety of apple.

Summary of key points

1. The null hypothesis H_0 is the hypothesis that we assume to be correct.
2. The alternative hypothesis H_1 tells us about the parameter if our assumption is shown to be wrong.
3. Hypothesis tests with alternative hypotheses in the form $H_1: \theta < \ldots$ and $H_1: \theta > \ldots$ are called one-tailed tests, where θ is the parameter from the distribution.
4. Hypothesis tests with an alternative hypothesis in the form $H_1: \theta \neq \ldots$ are called two-tailed tests, where θ is the parameter from the distribution.
5. A critical region is a region of the probability distribution which, if the test statistic falls within it, would cause you to reject the null hypothesis.
6. The critical value is the first value to fall inside of the critical region.
7. The actual significance level of a hypothesis test is the probability of incorrectly rejecting the null hypothesis.
8. For a two-tailed test, the critical region is split at either end of the distribution.
9. For a two-tailed test, halve the significance level at each end you are testing.

Review exercise 2

1 The random variable X has a continuous uniform distribution on $[a, b]$, where a and b are positive numbers. Given that $E(X) = 31$ and $Var(X) = 192$,

a find the value of a and the value of b

b find $P(X < 22 \mid 10 < X < 30)$

← Statistics 2 Section 5.1

(E) 2 The random variable X is uniformly distributed over the interval $[-1, 5]$.

a Sketch the probability density function $f(x)$ of X. **(2)**

Find:

b $E(X)$ **(2)**

c $Var(X)$ **(2)**

d $P(-0.3 < X < 3.3)$ **(2)**

← Statistics 2 Section 3.5

(E) 3 The continuous random variable X is uniformly distributed over the interval $[2, 6]$.

a Write down the probability density function $f(x)$. **(2)**

Find:

b $E(X)$ **(2)**

c $Var(X)$ **(2)**

d the cumulative distribution function of X, for all x **(2)**

e $P(2.3 < X < 3.4)$ **(2)**

← Statistics 2 Section 3.5

(E) 4 A string AB of length 5 cm is cut, in a random place C, into two pieces. The random variable X is the length of AC.

a Write down the name of the probability distribution of X and sketch the graph of its probability density function. **(3)**

b Find the values of $E(X)$ and $Var(X)$. **(4)**

c Find $P(X > 3)$. **(2)**

d Write down the probability that AC is exactly 3 cm long. **(1)**

← Statistics 2 Sections 3.5, 3.6

(E) 5 The continuous random variable X is uniformly distributed over the interval $\alpha < x < \beta$.

a Write down the probability density function of X, for all x. **(2)**

b Given that $E(X) = 2$ and $P(X < 3) = \frac{5}{8}$, find the value of α and the value of β. **(3)**

← Statistics 2 Sections 3.5, 3.6

(E) 6 A gardener has wire cutters and a piece of wire 150 cm long which has a ring attached at one end. The gardener cuts the wire, at a randomly chosen point, into two pieces. The length, in cm, of the piece of wire with the ring on it is represented by the random variable X.

Find:

a $E(X)$ **(2)**

b the standard deviation of X **(2)**

c the probability that the shorter piece of wire is at most 30 cm long. **(2)**

← Statistics 2 Sections 3.5, 3.6

7 The continuous random variable X is uniformly distributed over the interval $[a, b]$.

Given that $P(9 < X < 14) = 0.4$ and $E(X) = 12$,

a find the value of a and the value of b

b find the value of the constant k, such that $E(kX - 3) = 0$

c find the exact value of $E(X^2)$

d find $P(5X - b > a)$

← Statistics 2 Section 5.1

E 8 **a** Explain briefly what you understand by:

i a sampling frame

ii a statistic.

A random sample $X_1, X_2, \ldots, X_n$ is taken from a population with unknown mean μ.

b For each of the following, state whether or not it is a statistic. Give a reason for each of your answers.

i $\frac{X_1 + X_5 + X_n}{3}$

ii $\frac{\sum X^2}{n} - \mu$

← Statistics 2 Sections 1.1, 1.2, 1.3

9 A bag contains a large number of red, yellow and blue counters in the ratio 3 : 1 : 2.

Two counters are drawn at random.

a Write down all possible samples.

Five points are awarded when the counters are the same colour; one point is awarded when they are different colours.

b Find the sampling distribution for the number of points awarded.

c Estimate the total number of points if this experiment was carried out 18 times.

← Statistics 2 Section 6.1, 6.2, 6.3

10 A bag contains a large number of \$1, \$2 and \$5 coins in the ratio 3 : 3 : 4

A random sample of three coins is taken from the bag.

Find the sampling distribution of the median of these samples.

← Statistics 2 Section 6.1, 6.2, 6.3

11 **a** Explain what you understand by:

i a hypothesis test

ii a critical region.

During term time, incoming calls to a school are thought to occur at a rate of 0.45 per minute. To test this, the number of calls during a random 20-minute interval is recorded.

b Find the critical region for a two-tailed test of the hypothesis that the number of incoming calls occurs at a rate of 0.45 per 1-minute interval. The probability in each tail should be as close to 2.5% as possible.

c Write down the actual significance level of the above test.

In the school holidays, one call occurs in a 10-minute interval.

d Test, at the 5% level of significance, whether or not there is evidence that the rate of incoming calls is less during the school holidays than in term time.

← Statistics 2 Sections 1.1, 1.2, 1.3

12 Bacteria are randomly distributed in a river at a rate of 5 bacteria per litre of water. A new factory opens, and a scientist claims that it is polluting the river with bacteria. The scientist takes a sample of 0.5 litres of water from the river near the factory and it contains 7 bacteria. Stating your hypotheses clearly, test his claim at the 5% level of significance.

← Statistics 2 Sections 1.1, 1.2, 1.3

13 A manufacturer of coloured drawing pins introduces a purple pin that is to make up 15% of the total production. The pins are sold in boxes of 20.

a Find the critical region for a two-tailed test of the hypothesis that the probability that a pin chosen at random is purple is 0.15.
The probability in either tail should be as close to 2.5% as possible.

b Write down the actual significance level of the test.

A teacher buys a box of pins and discovers that it contains only one purple pin.

c Test, at the 5% level of significance, whether or not there is evidence that the probability of a pin chosen at random being purple is less than 0.15

← Statistics 2 Sections 1.1, 1.2, 1.3

E/P **14** A single observation is taken from a test statistic $X \sim \mathrm{B}(40, p)$.
Given that H_0: $p = 0.3$ and H_1: $p \neq 0.3$:

a find the critical region for the test using a 2.5% significance level. The probability in each tail should be as close as possible to 1.25% **(4)**

b state the probability of incorrectly rejecting the null hypothesis using this test. **(1)**

← Statistics 2 Section 7.2

E/P **15** A medicine company claims that 75% of patients suffering from depression recover when treated with a new medicine.

A random sample of 10 patients with depression is taken from a doctor's records.

a Write down a suitable distribution to model the number of patients in this sample who recover when treated with the new medicine. **(1)**

Given that the claim is correct,

b find the probability that the treatment will be successful for exactly 6 patients. **(2)**

The doctor believes that the claim is incorrect and the percentage who will recover is lower. From her records, she took a random sample of 20 patients who had been treated with the new drug. She found that 13 had recovered.

c Stating your hypotheses clearly, test, at the 5% level of significance, the doctor's belief. **(7)**

d From a sample of size 20, find the greatest number of patients who need to recover from the test in part **c** to be significant at the 1% level. **(3)**

← Statistics 2 Sections 6.3, 7.1, 7.2, 7.3

E/P **16** Dhriti grows tomatoes. Over a period of time, she has found that there is a probability 0.3 of a ripe tomato having a diameter greater than 4 cm. Dhriti wants to test whether a new fertiliser increases the size of the tomatoes. She takes a sample of 40 ripe tomatoes that have been treated with the new fertiliser.

a Write down suitable hypotheses for her test. **(1)**

b Using a 5% significance level, find the critical region for her test. **(4)**

c Write down the actual significance level of the test. **(1)**

Dhriti finds that 18 out of the 40 tomatoes have a diameter greater than 4 cm.

d Comment on Dhriti's observed value in light of your answer to part **b**. **(2)**

← Statistics 2 Sections 7.2, 7.3

E/P **17** Eva and Theo work in a microchip manufacturing plant. Theo claims that the chance of a chip being faulty is 1 in 2000. Eva suspects that this is an underestimate (a guess that it is smaller than the actual value) and that it is more likely that a randomly chosen chip is faulty.

Eva decided to test Theo's claim at the 5% significance level by sampling from a large batch of chips until she finds a faulty one.

a Find the critical region for Eva's test. **(5)**

b Eva finds the first faulty chip after selecting her 115th. Comment on this, taking your answer to part **a** into consideration. **(2)**

← Statistics 2 Section 4.4

Challenge

1 Let $X_1, X_2, X_3, \ldots X_n$ be identically distributed independent random variables, each with a continuous uniform distribution on [0, 1]. The random variable Y is defined as the maximum value taken by each X_i.

a Show that $\text{E}(Y) = \frac{n}{n+1}$

b Find an expression for the median of Y in terms of n.

If X and Y are independent continuous random variables with probability density functions f(x) and g(y) respectively, then the probability density function of $Z = X + Y$ is given by:

$$\text{h}(z) = \int_{-\infty}^{\infty} \text{f}(z-t)\text{g}(t)\,\text{d}t$$

Find and sketch the probability density function of:

c $X_1 + X_2$

d $X_1 + X_2 + X_3$.

← Statistics 2 Section 3.5

2 A test statistic has binomial distribution B(30, p). Given that H_0: $p = 0.65$ and H_1: $p < 0.65$,

a find the critical region for the test statistic such that the probability is as close as possible to 10%.

William takes two observations of the test statistic and finds that they both fall inside the critical region. He thus decides to reject H_0.

b Find the probability that William has incorrectly rejected H_0.

← Statistics 2 Section 5

Exam practice

Mathematics
International Advanced Subsidiary/ Advanced Level Statistics 2

Time: 1 hour 30 minutes
You must have: Mathematical Formulae and Statistical Tables, Calculator

---

1 A doctor conducts a random sample to find how many people in her town have flu. The probability that a person has flu is 0.15.

On Monday, she chooses to contact 12 people. Find the probability that on Monday:

a exactly 3 people have flu **(1)**

b no more than 5 people have flu. **(2)**

2 Let $X_1, X_2, \ldots, X_n$ be n independent observations from a population. Which of the following are statistics? Give a reason for your answer.

a $\dfrac{\overline{x} - \mu}{n}$ **(2)**

b $\max\{X_1, X_2, \ldots, X_n\}$ **(2)**

3 In a forest, the number of bear sightings occur at a rate of three per week.

a Find the probability that there are fewer than two bear sightings in a given week. **(2)**

b Find the probability that there are the expected number of bear sightings in a two week period. **(3)**

The forest is called 'bear safe' if, in 4 consecutive weeks, there are fewer than two bear sightings per week.

c Find the probability that the forest is called 'bear safe'. **(4)**

4 A random variable X has a probability density function given by:

$$f(x) = \begin{cases} \frac{1}{33}(ax + b) & 1 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Given that $E(X) = \frac{26}{11}$,

a show that $14a + 5b = 52$. **(4)**

b Find the values of a and b. **(4)**

c Show that the upper quartile is 3.08, correct to 2 decimal places. **(5)**

d State the mode of $f(x)$. **(1)**

5 The length L, in minutes, of a queue for a theme park ride can be modelled as a uniform distribution, $L \sim U[0, 80]$.

a Write down the full definition of the probability density function. **(1)**

Yifan joins the queue at 11:30 am. She has to meet her friend Daiyu at 12:45. The time taken for the ride is 3 minutes and it will take Yifan 2 minutes to walk to meet Daiyu from the end of the ride. She will need to leave the queue at 12:40 to meet Daiyu if she has not yet been on the ride.

b Find the probability that Yifan will need to leave the queue without going on the ride. **(3)**

Yifan decides to ride again later.

c Given that she has been in the queue for 30 minutes, find the probability that she will queue for at least a further 35 minutes before going on the ride. **(3)**

6 Typing errors occur in a newspaper at a rate of 4 every page.

a Calculate the probability that on two randomly selected pages,

i exactly 7 errors will occur

ii more than 7 errors will occur. **(4)**

Using a normal approximation, the probability that more than 40 errors will occur in n pages is 0.2268, correct to 4 significant figures.

b Find the value of n. **(9)**

7 Benoit planted 40 seeds in a tree nursery. He read in a journal that the probability of one of these seeds growing is 0.2. Fifteen of Benoit's seeds grew. This led Benoit to claim that the probability given by the journal was less than it should have been.

a Using a 1% level of significance, find the critical region for the test above. **(2)**

b Stating your hypothesis clearly, what conclusions can you make from your answer to part **a**? **(4)**

c What is the probability of Benoit incorrectly rejecting the null hypothesis? **(2)**

8 Brigitta has a fair spinner whose scores are $B = 1, 2, 3, 4$, and Roberto has a different fair spinner whose scores are $R = 1, 2, 3$. Both spinners are spun.

The statistic $X = (B - 1) \times (3 - R)$ is calculated.

a Show that $\mathrm{P}(X = 4) = \frac{1}{12}$ **(3)**

b Find the sampling distribution of X. **(4)**

c Write down the mode of X. **(1)**

9 **a** Write down two conditions under which the Poisson distribution may be used as an approximation to the binomial distribution. **(2)**

A machine that makes lightbulbs is known to produce 0.3% faulty lightbulbs. The machine breaks down and a new machine is installed. A random sample of 2000 bulbs is taken from those produced by the new machine and 12 light bulbs are found to be faulty.

b Using a suitable approximation, test, at the 5% level of significance, whether or not the proportion of faulty lightbulbs is higher with the new machine than with the old machine. State your hypotheses clearly. **(7)**

TOTAL FOR PAPER: 75 MARKS

Binomial Cumulative Distribution Function

The tabulated value is $P(X \leqslant x)$, where X has a binomial distribution with index n and parameter p.

$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 5, x = 0$	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
$n = 6, x = 0$	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
$n = 7, x = 0$	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
$n = 8, x = 0$	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
$n = 9, x = 0$	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
$n = 10, x = 0$	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990

$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 12, x = 0$	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
$n = 15, x = 0$	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 20, x = 0$	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 25, x = 0$	0.2774	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.6424	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000
2	0.8729	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000
3	0.9659	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001
4	0.9928	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005
5	0.9988	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020
6	0.9998	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073
7	1.0000	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216
8	1.0000	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539
9	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148
10	1.0000	1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122
11	1.0000	1.0000	0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450
12	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000
13	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550
14	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878
15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9971	0.9868	0.9560	0.8852
16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9957	0.9826	0.9461
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9942	0.9784
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9927
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9980
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 30, x = 0$	0.2146	0.0424	0.0076	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.5535	0.1837	0.0480	0.0105	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000
2	0.8122	0.4114	0.1514	0.0442	0.0106	0.0021	0.0003	0.0000	0.0000	0.0000
3	0.9392	0.6474	0.3217	0.1227	0.0374	0.0093	0.0019	0.0003	0.0000	0.0000
4	0.9844	0.8245	0.5245	0.2552	0.0979	0.0302	0.0075	0.0015	0.0002	0.0000
5	0.9967	0.9268	0.7106	0.4275	0.2026	0.0766	0.0233	0.0057	0.0011	0.0002
6	0.9994	0.9742	0.8474	0.6070	0.3481	0.1595	0.0586	0.0172	0.0040	0.0007
7	0.9999	0.9922	0.9302	0.7608	0.5143	0.2814	0.1238	0.0435	0.0121	0.0026
8	1.0000	0.9980	0.9722	0.8713	0.6736	0.4315	0.2247	0.0940	0.0312	0.0081
9	1.0000	0.9995	0.9903	0.9389	0.8034	0.5888	0.3575	0.1763	0.0694	0.0214
10	1.0000	0.9999	0.9971	0.9744	0.8943	0.7304	0.5078	0.2915	0.1350	0.0494
11	1.0000	1.0000	0.9992	0.9905	0.9493	0.8407	0.6548	0.4311	0.2327	0.1002
12	1.0000	1.0000	0.9998	0.9969	0.9784	0.9155	0.7802	0.5785	0.3592	0.1808
13	1.0000	1.0000	1.0000	0.9991	0.9918	0.9599	0.8737	0.7145	0.5025	0.2923
14	1.0000	1.0000	1.0000	0.9998	0.9973	0.9831	0.9348	0.8246	0.6448	0.4278
15	1.0000	1.0000	1.0000	0.9999	0.9992	0.9936	0.9699	0.9029	0.7691	0.5722
16	1.0000	1.0000	1.0000	1.0000	0.9998	0.9979	0.9876	0.9519	0.8644	0.7077
17	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9955	0.9788	0.9286	0.8192
18	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9917	0.9666	0.8998
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9971	0.9862	0.9506
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9950	0.9786
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9919
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9974
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 40, x = 0$	0.1285	0.0148	0.0015	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.3991	0.0805	0.0121	0.0015	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.6767	0.2228	0.0486	0.0079	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000
3	0.8619	0.4231	0.1302	0.0285	0.0047	0.0006	0.0001	0.0000	0.0000	0.0000
4	0.9520	0.6290	0.2633	0.0759	0.0160	0.0026	0.0003	0.0000	0.0000	0.0000
5	0.9861	0.7937	0.4325	0.1613	0.0433	0.0086	0.0013	0.0001	0.0000	0.0000
6	0.9966	0.9005	0.6067	0.2859	0.0962	0.0238	0.0044	0.0006	0.0001	0.0000
7	0.9993	0.9581	0.7559	0.4371	0.1820	0.0553	0.0124	0.0021	0.0002	0.0000
8	0.9999	0.9845	0.8646	0.5931	0.2998	0.1110	0.0303	0.0061	0.0009	0.0001
9	1.0000	0.9949	0.9328	0.7318	0.4395	0.1959	0.0644	0.0156	0.0027	0.0003
10	1.0000	0.9985	0.9701	0.8392	0.5839	0.3087	0.1215	0.0352	0.0074	0.0011
11	1.0000	0.9996	0.9880	0.9125	0.7151	0.4406	0.2053	0.0709	0.0179	0.0032
12	1.0000	0.9999	0.9957	0.9568	0.8209	0.5772	0.3143	0.1285	0.0386	0.0083
13	1.0000	1.0000	0.9986	0.9806	0.8968	0.7032	0.4408	0.2112	0.0751	0.0192
14	1.0000	1.0000	0.9996	0.9921	0.9456	0.8074	0.5721	0.3174	0.1326	0.0403
15	1.0000	1.0000	0.9999	0.9971	0.9738	0.8849	0.6946	0.4402	0.2142	0.0769
16	1.0000	1.0000	1.0000	0.9990	0.9884	0.9367	0.7978	0.5681	0.3185	0.1341
17	1.0000	1.0000	1.0000	0.9997	0.9953	0.9680	0.8761	0.6885	0.4391	0.2148
18	1.0000	1.0000	1.0000	0.9999	0.9983	0.9852	0.9301	0.7911	0.5651	0.3179
19	1.0000	1.0000	1.0000	1.0000	0.9994	0.9937	0.9637	0.8702	0.6844	0.4373
20	1.0000	1.0000	1.0000	1.0000	0.9998	0.9976	0.9827	0.9256	0.7870	0.5627
21	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9925	0.9608	0.8669	0.6821
22	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9970	0.9811	0.9233	0.7852
23	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989	0.9917	0.9595	0.8659
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9804	0.9231
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9988	0.9914	0.9597
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9966	0.9808
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9988	0.9917
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9968
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$p =$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 50, x = 0$	0.0769	0.0052	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.2794	0.0338	0.0029	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.5405	0.1117	0.0142	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.7604	0.2503	0.0460	0.0057	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.8964	0.4312	0.1121	0.0185	0.0021	0.0002	0.0000	0.0000	0.0000	0.0000
5	0.9622	0.6161	0.2194	0.0480	0.0070	0.0007	0.0001	0.0000	0.0000	0.0000
6	0.9882	0.7702	0.3613	0.1034	0.0194	0.0025	0.0002	0.0000	0.0000	0.0000
7	0.9968	0.8779	0.5188	0.1904	0.0453	0.0073	0.0008	0.0001	0.0000	0.0000
8	0.9992	0.9421	0.6681	0.3073	0.0916	0.0183	0.0025	0.0002	0.0000	0.0000
9	0.9998	0.9755	0.7911	0.4437	0.1637	0.0402	0.0067	0.0008	0.0001	0.0000
10	1.0000	0.9906	0.8801	0.5836	0.2622	0.0789	0.0160	0.0022	0.0002	0.0000
11	1.0000	0.9968	0.9372	0.7107	0.3816	0.1390	0.0342	0.0057	0.0006	0.0000
12	1.0000	0.9990	0.9699	0.8139	0.5110	0.2229	0.0661	0.0133	0.0018	0.0002
13	1.0000	0.9997	0.9868	0.8894	0.6370	0.3279	0.1163	0.0280	0.0045	0.0005
14	1.0000	0.9999	0.9947	0.9393	0.7481	0.4468	0.1878	0.0540	0.0104	0.0013
15	1.0000	1.0000	0.9981	0.9692	0.8369	0.5692	0.2801	0.0955	0.0220	0.0033
16	1.0000	1.0000	0.9993	0.9856	0.9017	0.6839	0.3889	0.1561	0.0427	0.0077
17	1.0000	1.0000	0.9998	0.9937	0.9449	0.7822	0.5060	0.2369	0.0765	0.0164
18	1.0000	1.0000	0.9999	0.9975	0.9713	0.8594	0.6216	0.3356	0.1273	0.0325
19	1.0000	1.0000	1.0000	0.9991	0.9861	0.9152	0.7264	0.4465	0.1974	0.0595
20	1.0000	1.0000	1.0000	0.9997	0.9937	0.9522	0.8139	0.5610	0.2862	0.1013
21	1.0000	1.0000	1.0000	0.9999	0.9974	0.9749	0.8813	0.6701	0.3900	0.1611
22	1.0000	1.0000	1.0000	1.0000	0.9990	0.9877	0.9290	0.7660	0.5019	0.2399
23	1.0000	1.0000	1.0000	1.0000	0.9996	0.9944	0.9604	0.8438	0.6134	0.3359
24	1.0000	1.0000	1.0000	1.0000	0.9999	0.9976	0.9793	0.9022	0.7160	0.4439
25	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9900	0.9427	0.8034	0.5561
26	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9955	0.9686	0.8721	0.6641
27	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9840	0.9220	0.7601
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9924	0.9556	0.8389
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9966	0.9765	0.8987
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9986	0.9884	0.9405
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9947	0.9675
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9978	0.9836
33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991	0.9923
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9967
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9987
36	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
37	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
38	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Poisson Cumulative Distribution Function

The tabulated value is $P(X \leqslant x)$, where X has a Poisson distribution with parameter λ.

$\lambda =$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$x = 0$	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.9098	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9856	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9982	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9998	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	1.0000	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	1.0000	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8	1.0000	1.0000	1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\lambda =$	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
$x = 0$	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671
6	0.6860	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301
7	0.8095	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202
8	0.8944	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328
9	0.9462	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579
10	0.9747	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830
11	0.9890	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968
12	0.9955	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916
13	0.9983	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645
14	0.9994	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165
15	0.9998	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513
16	0.9999	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730
17	1.0000	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857
18	1.0000	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928
19	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965
20	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997

GLOSSARY

acceptance region when doing a hypothesis test, if the test statistic is in the acceptance region, we do not reject the null hypothesis

alternative hypothesis when doing a hypothesis test, the alternative hypothesis tells you about the parameter if your assumption is shown to be wrong

approximation a non-exact answer, usually found because the exact calculation is too time-consuming

assumption something which you believe to be true

Bernoulli trial a single trial where there are two outcomes: success and failure

binomial distribution a distribution where each trial has two outcomes (success and failure), the trials are independent and have a fixed probability of success

boundary value in a function, it is the value of the function at the beginning or end of the domain

census observes or measures every member of a population

component one part of a larger whole, especially part of a machine or system

condition something that must happen before something else can happen

consecutive occurring one after another

constant staying at the same level; a term that has no **variable**, e.g. the constant term in the expression $x^2 + 3x - 6$ is -6

continuity correction an adjustment made when you use a continuous probability distribution to approximate a discrete probability distribution

continuous data that can take any value in a given range

continuous random variable a **variable** that can take any value in a given range

continuous uniform distribution a continuous probability distribution which has a constant probability function

critical region when doing a hypothesis test, if the test statistic is in the critical region, we reject the null hypothesis

critical value the first value to fall inside the critical region

cumulative to increase in value by successive additions

cumulative probability function a function which will calculate the probability of being below a certain value of a random variable

discrete data that can take only certain values in a given range

discrete random variable only takes values on a discrete scale

expectation the mean of a random variable

inclusive including all the values mentioned

indefinite integral an integral without upper and lower limits

independent when one event has no effect on another for the

index for the binomial, it is the number of trials

infinitely with no limit

interval a period of time or space between two events; a range of values

likelihood how likely it is that something will happen

lower quartile a value which has 25% of the probability below it

mean the sum/integral of all the observations divided by the total number of observations

median a value which has 50% of the probability below it and 50% above it

mode the highest value for the probability density function

model to use **assumptions** to make a real-world example easier to understand

null hypothesis when doing a hypothesis test, the null hypothesis is the assumption about the parameter you are going to make

outcome the result of an experiment

parameter a value (which is unknown) that describes a feature of a population

piecewise function a function which is defined by several sub-functions, each applying to a certain interval of the domain

population parameter see **parameter**

probability a measure showing how likely it is that something will happen

probability density function a function that describes the relative **likelihood** for the random variable to take on a given value

proportional increasing or decreasing in amount or degree in relation to changes in something else

random chosen by chance; without any regular pattern

sample a small representative set from a population

sampling distribution a distribution of the values that a statistic can take, along with their probabilities

sampling frame a list/register of sampling units

sampling unit the individual units of a population

satisfy to meet the requirements

significance level the probability of rejecting the null hypothesis in a hypothesis test

singly at a given point in an interval, only one event can happen

statistic a value which is a function of only the data from a sample. It cannot contain any parameters

symmetrical made of two exact matching halves

test statistic a value from observations which is used in a hypothesis test

threshold the point when something starts to happen or change. In a hypothesis test, it is called a significance level

trial an experiment which an be repeated many times with a known set of outcomes

uniformly the same in all parts

upper quartile a value which has 75% of the probability below it

variable able to be changed. Represented by a symbol (X, Y, A, B etc.) and able to take on any of a specified set of values

variance a measure of how spread out the distribution is

ANSWERS

CHAPTER 1

Prior knowledge check

1 a $\frac{1}{8}$ b $\frac{1}{8}$ c $\frac{3}{8}$ d $\frac{1}{2}$
2 a $\frac{4}{36}$ b $\frac{18}{36}$ c $\frac{18}{36}$ d $\frac{12}{36}$ e $\frac{15}{36}$

Exercise 1A

1 a 0.273 b 0.0683 c 0.195
2 a 0.00670 b 0.214 c 0.00178
3 a $X \sim B(20, 0.01)$, $n = 20$, $p = 0.01$
Assume bolts being defective are independent of each other.
b $X \sim B(6, 0.52)$, $n = 6$, $p = 0.52$
Assume the lights operate independently and the time lights are on/off is constant.
c $X \sim B(30, \frac{1}{8})$, $n = 30$, $p = \frac{1}{8}$
Assume serves are independent and probability of an ace is constant.
4 a $X \sim B(14, 0.15)$ is OK if we assume the children in the class being Rh^- is independent from child to child (so no siblings/twins).
b This is not binomial since the number of flips is not fixed. The probability of a head at each flip is constant ($p = 0.5$) but there is no value of n.
c Assuming the colours of the cars are independent (which should be reasonable).
X = number of red cars out of 15
$X \sim B(15, 0.12)$
5 a 0.358 b 0.189
6 a The random variable can take two values, faulty or not faulty.
There are a fixed number of trials, 10, and fixed probability of success: 0.08.
Assuming each member in the sample is independent, a suitable model is $X \sim B(10, 0.08)$
b 0.00522
7 a Assumptions: There is a fixed sample size, there are only two outcomes for the genetic marker (i.e. fully present or not present), there is a fixed probability of people having the marker.
b 0.0108
8 a The random variable can take two values, 6 or not 6. There are a fixed number of trials (15) and a fixed probability of success (0.3), Each roll of the dice is independent. A suitable distribution is $X \sim B(15, 0.3)$
b 0.219 c 0.127

Exercise 1B

1 a 0.9804 b 0.7382 c 0.5638 d 0.3020
2 a 0.9468 b 0.5834 c 0.1272 d 0.5989
3 a 0.8702 b 0.4319 c 0.5246 d 0.5329
4 a 0.9816 b 0.0004 c 0.0358
5 a 0.0039 b 0.9648 c 0.3633
6 a 0.2252 b 0.4613 c 0.7073
7 a $k = 13$ b $r = 28$
8 a $k = 1$ b $r = 9$ c 0.9801

9 a $X \sim B(10, 0.30)$ Assumptions: The random variable can take two values (listen or don't listen), there are a fixed number of trials (10) and a fixed probability of success (0.3), each member in the sample is independent.
b 0.1503 c $s = 8$
10 a 0.2794 b 0.0378 c $d = 5$

Exercise 1C

1 a 8.4 b 2.52
2 a 8 b 0.1239 c 0.3154
3 0.4 or 0.6
4 0.2 or 0.8
5 $n = 12$, $p = 0.4$
6 a 0.3 b 0.1643
7 a i 0.1536 ii 0.9051
b i 120 ii 27.3
8 a 0.5772 b Mean = 69.26 Variance = 29.28
9 a 0.3 b $E(X) = 1.5$, $Var(X) = 1.05$
10 a Mean = 1, Variance = 0.8
b 0.2
c 164, 205, 102, 26, 3, 0
The values support the student's suggestion that the data can be modelled by a binomial distribution
d Variance = $5 \times 0.2 \times 0.8 = 0.8$
The calculated variance matches the observed variance of the data and supports the use of a binomial distribution

Challenge

a $P(X = 0) = \binom{3}{0} \times p^0 \times (1 - p)^3 = (1 - p)^3$

$P(X = 1) = \binom{3}{1} \times p^1 \times (1 - p)^2 = 3p(1 - p)^2$

$P(X = 2) = \binom{3}{2} \times p^2 \times (1 - p)^1 = 3p^2(1 - p)$

$P(X = 3) = \binom{3}{3} \times p^3 \times (1 - p) = p^3$

$E(X) = \sum XP(X = x)$
$= (0 \times (1 - p)^3) + (1 \times 3p(1 - p)^2) + (2 \times 3p^2(1 - p))$
$+ (3 \times p^3)$
$= 3p - 6p^2 + 3p^3 + 6p^2 - 6p^3 + 3p^3 = 3p$

b $E(X^2) = (0^2 \times (1 - p)^3) + (1^2 \times 3p(1 - p)^2)$
$+ (2^2 \times 3p^2(1 - p)) + (3^2 \times p^3)$
$= 3p - 6p^2 + 3p^3 + 12p^2 - 12p^3 + 9p^3 = 3p + 6p^2$
$Var(X) = E(X^2) - E^2(X) = 3p + 6p^2 - (3p)^2 = 3p(1 - p)$

Chapter review 1

1 a 0.114 b 0.0005799
c 0.9373
2 a 0.0439 or $\frac{32}{729}$ b 0.273
3 a 0.014 (3 d.p.) b 0.747 (3 d.p.)
4 a 1 There are n independent trials.
2 n is a fixed number.
3 The outcome of each trial is success or failure.
4 The probability of success at each trial is constant.
5 The outcome of any trial is independant of any other trial.
b 0.0861 c $n = 90$

5 **a** 0.000977 **b** 0.0547
6 **a** 0.0531 **b** 0.243
7 **a** $X \sim B(10, 0.15)$
b 0.0099 **c** 0.2759
8 **a** 0.8692 **b** 0.0727
9 **a** 0.8725 **b** 0.01027 **c** 0.0002407
10 **a** **i** 0.2627 **ii** 0.0582
b 0.2560
11 **a** 0.0162 **b** Mean = 10, Variance = 9.8
12 **a** 0.0230 **b** Mean = 3, Variance = 2.925
13 $n = 5$

Challenge
0.001244

CHAPTER 2

Prior knowledge check

1 **a** $E(X) = 3.8$ **b** $E(X^2) = 18$ **c** $Var(X) = 3.56$
2 **a** 0.0168 **b** 0.001 **c** 0.3972

Exercise 2A

1 **a** 0.2138 **b** 0.7127 **c** 0.4703
2 **a** 0.1733 **b** 0.8153 **c** 0.7531
3 **a** 0.1323 **b** 0.3954 **c** 0.5429
4 **a** 0.3626 **b** 0.5683 **c** 0.1950
5 $\lambda = 3$
6 $\lambda = 6$

Exercise 2B

1 **a** 0.2017 **b** 0.4711 **c** 0.7211
2 **a** 0.7798 **b** 0.6615 **c** 0.3035
3 **a** 0.8641 **b** 0.6139 **c** 0.5368
4 **a** 0.4679 **b** 0.3606 **c** 0.8200
5 **a** 6 **b** 9 **c** 5 **d** 5
6 **a** 5 **b** 2 **c** All values of $c > 6$
d All values of $d > 8$

Exercise 2C

1 **a** **i** 0.1680 **ii** 0.0839
b **i** 0.1606 **ii** 0.2851
2 **a** (1) Weeds grow independently
(2) Weeds grow at a constant rate/unit of area
b 0.1088 **c** 0.2084
3 **a** $X \sim Po(2.5)$
b (1) Faults occur independently
(2) Faults occur at a constant rate
c 0.2565 **d** 0.7586 **e** 0.8699
4 **a** **i** 0.1755 **ii** 0.0681
b **i** 0.7586 **ii** 0.8622
5 **a** 0.0839 **b** 0.1512
6 **a** 0.1247 **b** 0.0137
7 **a** **i** 0.1653 **ii** 0.1607 **iii** 0.2694
b 0.4202
8 **a** **i** 0.1336 **ii** 0.4562
b 0.7135
9 **a** **i** 0.5276 **ii** 0.1329
b 0.5276 – breakdowns occur independently of each other
10 **a** 0.1255 **b** 0.1512 **c** 0.1670
11 **a** 0.1247 **b** 0.0260
12 **a** 0.2650 **b** 11 buses
13 **a** 0.2971
b $Y \sim Po(3)$. $P(X > 8) = 1 - P(X \leqslant 8) = 1 - 0.9962 = 0.0038 = 0.38\%$
c 10
14 **a** 0.8088 **b** 0.1847 **c** 14

Exercise 2D

1 **a** 0.1606 **b** 0.7440 **c** 0.7149
2 **a** 0.1465 **b** 0.2414 **c** 0.2236
3 **a** 0.0474 **b** 0.6159 **c** 0.3099 **d** 0.2851
4 **a** 0.5049 **b** 0.3134
5 **a** **i** 0.1213 **ii** 0.7166
b Events occur at a constant average rate – the mean number of an interval is proportional to the length of the interval.
6 **a** 0.2238 **b** 0.2707 **c** 0.4579
7 **a** 0.6988 **b** 0.3153 **c** 0.1607
8 **a** 0.2090 **b** 0.3374 **c** 0.4457
9 **a** 0.0158 **b** 0.7534 **c** 0.2417
10 **a** 0.2639 **b** 0.7657 **c** 0.0754

Challenge

a $Q \sim Po(\lambda + \mu)$

$$P(Q = 0) = \frac{(e^{-(\lambda+\mu)} \times (\lambda + \mu)^0)}{0!}$$

$(\lambda + \mu)^0 = 1$ and $(0!) = 1$

therefore $P(Q = 0) = e^{(-\lambda + \mu)}$

$Q \sim Po(\lambda + \mu)$

b $$P(Q = 1) = \frac{(e^{-(\lambda+\mu)} \times (\lambda + \mu)^1)}{1!}$$

$(\lambda + \mu)^1 = (\lambda + \mu)$ and $1! = 1$

therefore $P(Q = 1) = e^{-(\lambda + \mu)} \times (\lambda + \mu)$

Exercise 2E

1 **a** Mean = 1.43, Variance = 1.4251
b Mean ≈ variance
c Using $\lambda = 1.4$, P = 0.1128
2 **a** Mean = 3.64, Variance = 3.5604
b Mean ≈ variance
c Using $\lambda = 3.6$, $P(X \leqslant 2) = 0.3027$
d From the table relative frequency of obtaining no more than 2 cars per period is 0.29. Answer from **c** is very close to recorded value.
3 **a** Mean = 2.867, Variance = 2.897
b Mean ≈ variance
c Because the observed frequency for 8 or more flaws was 0.
d 99 (using $\lambda = 2.9$)

Challenge

Proof outline:

$$E(X) = \sum_{i=0}^{\infty} i \times P(X = i) = \sum_{i=0}^{\infty} \frac{i \times e^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=1}^{\infty} \frac{i \times \lambda^i}{(i-1)!}$$

$$= e^{-\lambda}\sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{i!} = e^{-\lambda} \times \lambda\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \times \lambda \times e^{\lambda} = \lambda$$

$Var(X) = E(X^2) - E^2(X)$

Using a similar approach to above to gain: $E(X^2) = \lambda^2 + \lambda$

$Var(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda$

Chapter review 2

1 **a** 0.4966 **b** 0.2700 **c** 0.2376
2 **a** (1) Mistakes occur independently
(2) Mistakes occur at a constant rate
b 0.3425 **c** 0.1689
3 $\lambda = 5$
4 **a** (1) Emails arrive independently
(2) Emails arrive at a constant rate

b **i** 0.1377
ii 0.2560

5 **a** n is large, p is small
b 0.4253 **c** 0.4335 **d** 1.93%

6 $\lambda = 7$

7 0.2022

8 **a** **i** 0.1336 **ii** 0.4562
b 0.2084 **c** 0.2992

9 **a** $X \sim \text{Po}(6)$, properties are sold independently and at a constant rate
b 0.1606
c 0.1090 or 0.1091 (using unrounded answer for part **a**)

10 **a** 0.3848 **b** 0.1804 **c** 0.0440

11 **a** 0.3285 **b** 0.1042 **c** 0.3134

12 **a** 0.1321 **b** 0.7135 **c** 0.3191

13 **a** 0.1141 **b** 0.0103

14 **a** $X \sim \text{Po}(4)$ Website visits occur independently of each other and at a constant average rate.
b 0.0298 **c** 0.2834

15 **a** Mean = 2.86, Variance = 2.867
b Mean ≈ variance

Challenge

a $\frac{63}{256}$ or 0.2461 (4 d.p.)

b $\frac{7}{128}$ or 0.0547 (4 d.p.)

CHAPTER 3

Prior knowledge check

1 **a** 0.124 (3 s.f.) **b** 0.584 (3 s.f.) **c** 0.869 (3 s.f.)

2 **a** 0.0.211 **b** 0.599

For Chapter 3, student answers may differ slightly from those shown here when calculators are used rather than table values.

Exercise 3A

1 **a** **i** 0.1781 **ii** 0.1183
b **i** 0.1755 **ii** 0.1247

2 **a** **i** 0.1628 **ii** 0.1458
b **i** 0.1606 **ii** 0.1512

3 **a** **i** 0.1963 **ii** 0.2351
b **i** 0.1954 **ii** 0.2381

4 **a** 0.1075 **b** 0.1074
c The two values are similar, so a Poisson distribution is a good approximation in this case.

5 **a** 0.6472 **b** 0.2240

6 **a** 0.0984 **b** 0.8743

7 **a** 0.1422 **b** 0.3782

8 **a** 0.1991 **b** 0.6472

9 **a** $X \sim \text{B}(10, 0.05)$ **b** 0.0105 **c** 0.5298

10 0.3498

11 **a** $X \sim \text{B}(1200, 0.005)$
b Mean = 6, Variance = 5.97
c 0.2851

12 **a** 0.2378 **b** 0.0315

13 **a** 0.7350 **b** 0.2788

Challenge

Consider $\text{P}(X = 0 \mid X \sim \text{B}(n, p)) = (1 - p)^n$

Consider $\text{P}(X = 0 \mid X \approx \text{Po}(np)) = \text{e}^{-np}$

It will overestimate when $\text{e}^{-np} > (1 - p)^n$

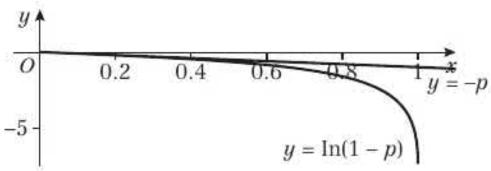
Taking logs:

$-np > n \ln(1 - p)$

Leading to:

$-p > \ln(1 - p)$

It can now be shown graphically.



Exercise 3B

1 **a** **i** Yes, n is large (> 50) and p is close to 0.5
ii $X \sim \text{N}(72, 5.37^2)$
b **i** No, n is not large enough (< 50).
c **i** Yes, n is large (> 50) and p is close to 0.5
ii $X \sim \text{N}(130, 7.90^2)$
d **i** No, p is too far from 0.5.
e **i** Yes, n is large (> 50) and p is close to 0.5
ii $X \sim \text{N}(192, 9.99^2)$
f **i** Yes, n is large (> 50) and p is close to 0.5
ii $X \sim \text{N}(580, 15.6^2)$

2 **a** 0.1253 **b** 0.0946 **c** 0.6723

3 **a** 0.0097 **b** 0.5596 **c** 0.0559

4 **a** 0.6203 **b** 0.4540 **c** 0.0102

5 0.006

6 0.3767

7 **a** n large, p close to 0.5
b 0.1593 **c** 0.5772 **d** 115

8 **a** 0.6277 **b** 0.8457

9 **a** 0.0786 **b** 0.26%

Exercise 3C

1 **a** 0.041-0.042 **b** 0.0069 **c** 0.683-0.685

2 **a** 0.206 **b** 0.251 **c** 0.456-0.457

3 **a** 0.626-0.627 **b** 0.0480 **c** 0.480-0.481

4 **a** 0.059-0.060 **b** 0.825-0.826 **c** 0.311-0.312

5 **a** 1 day **b** 4 or 5 days **c** 8 days

Chapter review 3

1 **a** n is large and p is close to 0.5
b $\mu = 40$, $\sigma^2 = 24$
c 0.0262

2 **a** 0.0147
b n is large and p is close to 0.5; $\mu = 55.2$, $\sigma = 5.46$
c 0.68%

3 **a** n is large and p is close to 0.5
b 0.5232 **c** 166

4 0.6339

5 **a** 0.5914 **b** 0.0197
c Assuming the claim is correct, there would be a less than 2% chance that 95 seedlings produce apples within 3 years. Therefore it is unlikely that the claim is correct.

6 **a** 0.5801 **b** 0.0594
c Assuming the claim is true, there is a less than 6% chance that 170 or more people would be cured out of 300, so it is likely that the herbalist has under-stated the actual cure rate.

7 **a** 0.8159 **b** 0.135-0.136

8 0.0778

9 **a** 0.0262 **b** 0.2149

10 **a** 0.1849 **b** 0.8576 **c** 0.946

11 **a** 0.104 **b** 0.8153

Challenge

a 0.2865
b 0.2873
c 0.2630
d Binomial to Poisson 0.279%
Binomial to normal 8.18%
Poisson is the more accurate approximation.

CHAPTER 4

Prior knowledge check

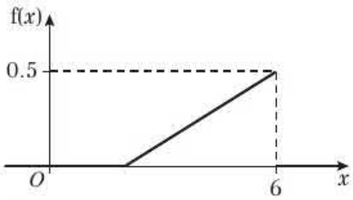
1 a $k = \frac{1}{20}$ **b** $P(X \geqslant 5) = \frac{7}{10}$ **c** $E(X) = 5$
2 a 4 **b** 3 **c** 48
3 $a = 6$

Exercise 4A

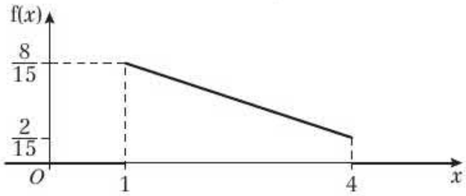
1 a There are negative values for $f(x)$ when $x < 0$ so this is not a probability density function.
b Area of $8\frac{2}{3}$ not equal to 1 therefore it is not a probability density function.
c There are negative values for $f(x)$ so this is not a probability density function.

2 $k = \frac{3}{50}$

3 a



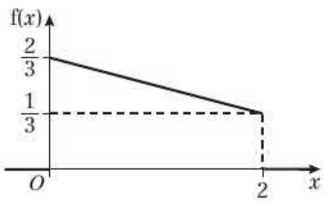
b



4 a $k = \frac{1}{4}$ **b** $k = \frac{3}{27} = \frac{1}{9}$ **c** $k = \frac{1}{6}$

5 a $k = \frac{1}{6}$

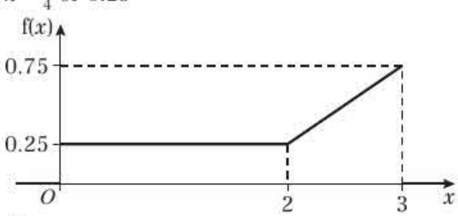
b



c $\frac{5}{12}$

6 a $k = \frac{3}{4}$ **b** $\frac{5}{16}$
7 a $k = \frac{4}{255}$ **b** $\frac{1}{17}$
8 a $k = \frac{1}{4}$ or 0.25

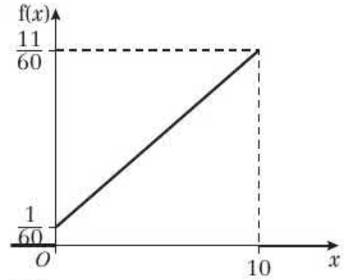
b



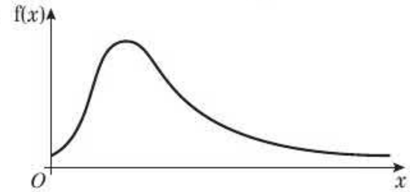
c $\frac{9}{64}$

9 a $\frac{1}{96}$

b

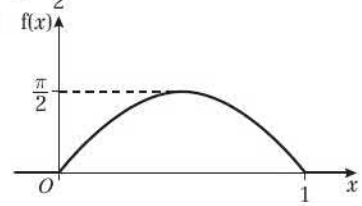


c



10 a $k = \dfrac{6}{11}$ **b** 0.295
11 a $k = \dfrac{1}{\ln 5}$ **b** $\dfrac{\ln 2}{\ln 5}$
12 a $k = \dfrac{1}{\ln 6}$ **b** 0.285
13 a $k = \dfrac{\pi}{2}$

b



c $\frac{1}{4}$

Challenge

a $k = 2$
b i $\frac{8}{9}$ **ii** $\frac{1}{400}$
c $p = 2.5$

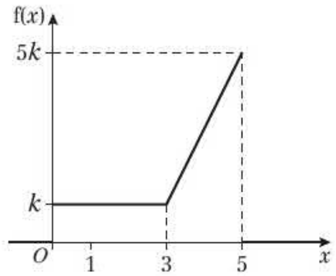
Exercise 4B

1
$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^3}{8} & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

2
$$F(x) = \begin{cases} 0 & x < 1 \\ x - \dfrac{x^2}{8} - \dfrac{7}{8} & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$$

3
$$F(x) = \begin{cases} 0 & x \leqslant 0 \\ \dfrac{x^2}{18} & 0 < x < 3 \\ \dfrac{2x}{3} - \dfrac{x^2}{18} - 1 & 3 \leqslant x \leqslant 6 \\ 1 & x > 6 \end{cases}$$

4 a



b $k = \frac{1}{9}$

c $$\text{F}(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{9} & 0 \leqslant x < 3 \\ \dfrac{x^2}{9} - \dfrac{5x}{9} + 1 & 3 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$

5 $$\text{f}(x) = \begin{cases} \dfrac{2x}{5} & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

6 a 0.75 **b** 0.75 **c** 0.5

7 $\frac{1}{6}2^p + q = 0$ (1) and $\frac{1}{6}4^p + q = 1$ (2)

(2) – (1): $\frac{1}{6}4^p - \frac{1}{6}2^p = 1 \Rightarrow 2^{2p} - 2^p = 6$

Let $y = 2^p$, then $y^2 - y - 6 = 0 \Rightarrow (y - 3)(y + 2) = 0$

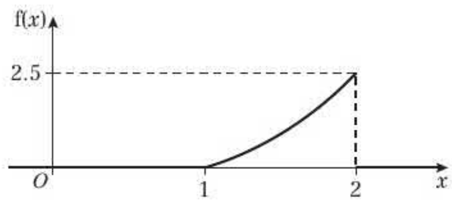
Taking the positive value, $y = 3 \Rightarrow 2^p = 3$

$p = \dfrac{\ln 3}{\ln 2}$

From (1), $q = -\frac{1}{2}$

8 a $$\text{f}(x) = \begin{cases} \dfrac{3}{2}x^2 - 2x + \dfrac{1}{2} & 1 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

b



c $\frac{3}{16}$

9 a $\int_0^2 k(4 - x^2)\,dx = 1 = \left[k\left(4x - \dfrac{x^3}{3}\right)\right]_0^2$

$\Rightarrow \dfrac{16k}{3} = 1 \Rightarrow k = \frac{3}{16}$

b $$\text{F}(x) = \begin{cases} 0 & x < 0 \\ \dfrac{3}{16}\left(4x - \dfrac{x^3}{3}\right) & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$$

c 0.007 (1 s.f.)

10 a $k = 49$ **b** 0.25

11 $$\text{F}(x) = \begin{cases} 0 & x < 1 \\ \dfrac{\ln x}{\ln 7} & 1 \leqslant x \leqslant 7 \\ 1 & x > 7 \end{cases}$$

12 $$\text{F}(x) = \begin{cases} 0 & x < 0 \\ \sin(\pi x) & 0 \leqslant x \leqslant \frac{1}{2} \\ 1 & x > \frac{1}{2} \end{cases}$$

13 a $k = \dfrac{1}{2 + \ln 3}$

b $$\text{f}(x) = \begin{cases} \dfrac{1}{2 + \ln 3}\left(1 + \dfrac{1}{x}\right) & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Challenge

a $$\text{F}(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-1.25t} & t \geqslant 0 \end{cases}$$

b 0.2044 (4 d.p.) **c** 0.0235 (4 d.p.)

Exercise 4C

1 a $k = \frac{3}{8}$ **b** $\frac{3}{2}$ **c** $\frac{3}{20}$

2 a 2.25 **b** 0.3375 **c** 0.581 (3 s.f.)

3 a $\frac{8}{3}$ **b** $\frac{8}{9}$ **c** 0.943 (3 s.f.)

d 0.556 (3 s.f.) **e** 8 **f** $\frac{14}{3}$

4 a $k = 2$ **b** $\frac{1}{3}$

c $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$

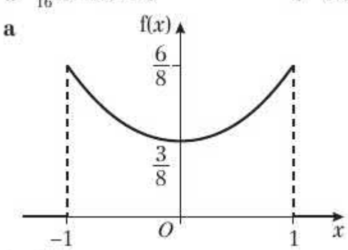
$= \int_0^1 2x^2(1 - x)\,dx - \left(\frac{1}{3}\right)^2$

$= \left[\dfrac{2x^3}{3} - \dfrac{2x^4}{4}\right]_0^1 - \frac{1}{9} = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$

d $\frac{4}{9}$

5 a $\frac{5}{16}$ or 0.3125 **b** 0.6 or $\frac{3}{5}$

6 a



b 0

c $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$

$= \int_{-1}^1 \frac{3}{8}x^2(1 + x^2)\,dx - 0^2$

$= \frac{3}{8}\left[\dfrac{x^3}{3} + \dfrac{x^5}{5}\right]_{-1}^1$

$= \frac{3}{8}\left(\frac{1}{3} + \frac{1}{5} - \left(-\frac{1}{3}\right) - \left(-\frac{1}{5}\right)\right) = 0.4$

d 0.538 (3 s.f.)

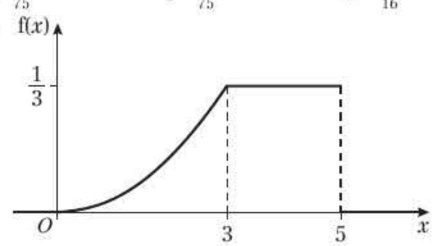
7 a $k = \frac{1}{4}$

b $\text{E}(T) = \frac{1}{4}\int_0^2 t^4\,dt = \frac{1}{4}\left[\dfrac{t^5}{5}\right]_0^2 = \frac{1}{4} \times \frac{32}{5} = 1.6$

c 6.2

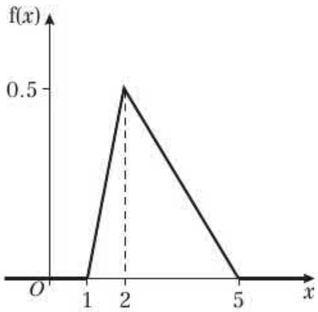
d $\frac{8}{75}$ **e** $\frac{32}{75}$ **f** $\frac{1}{16}$

8 a



b 3.417 **c** 1.0152 **d** 1.01

9 a



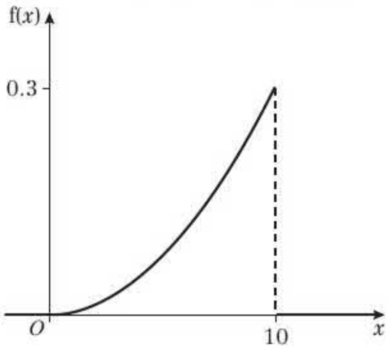
b $\frac{8}{3}$ **c** $\frac{13}{18}$

10 a $\int_0^{10} kt^2\,dt = 1 = \left[k\frac{t^3}{3}\right]_0^{10} = \frac{1000k}{3}$

$\Rightarrow k = \frac{3}{1000} = 0.003$

b 7.5 **c** 3.75 **d** 0.386

e



11 a $\frac{3}{4}$ **b** $\frac{4}{5}$ **c** $\frac{19}{80}$

12 a $\frac{50}{9}$

b $E(X^3) = \int_0^{10} \frac{x^4}{50}\,dx = \left[\frac{1}{250}x^5\right]_0^{10} = 400$

13 a $\frac{1}{\ln 3}$ **b** $\frac{2}{\ln 3}$ **c** 0.3268

14 a $1 = \int_1^2 \frac{c}{x(3-x)}\,dx = \frac{c}{3}\int_1^2 \left(\frac{1}{x} + \frac{1}{3-x}\right)dx$

$= \frac{c}{3}[\ln x - \ln(3-x)]_1^2 = \frac{c}{3}(2\ln 2) = \frac{c\ln 4}{3}$

$\Rightarrow c = \frac{3}{\ln 4}$

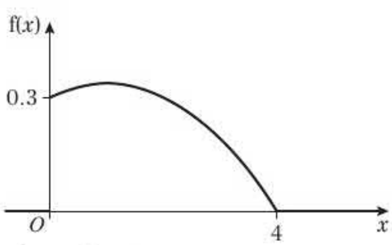
b $E(X) = 1.5$, $Var(X) = 0.0860$

Challenge

$-\frac{1}{2}$

Exercise 4D

1 a



b The mode is 1.

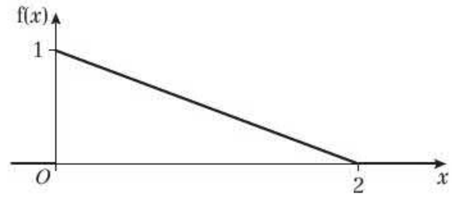
2 a $F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{16}x^2 & 0 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$

b 2.83

3 a Median = 1.732 since −1.732 is not in the range.

b $Q_1 = 1.225$, $Q_3 = 2.134$, IQR = 2.134 − 1.225 = 0.909

4 a



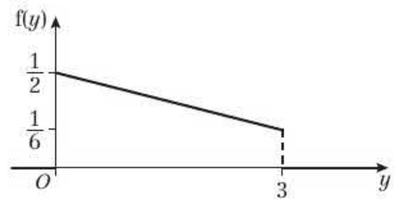
b 0

c $F(x) = \begin{cases} 0 & x < 0 \\ x - \frac{1}{4}x^2 & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

d Median = $2 - \sqrt{2} = 0.586$ (3 s.f.) as $2 + \sqrt{2}$ is not in range.

e 1

5 a

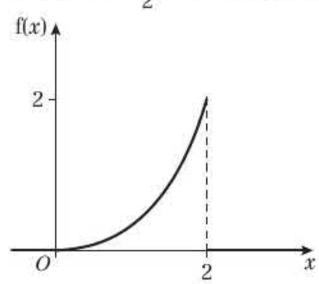


b 0

c $F(y) = \begin{cases} 0 & y < 0 \\ \frac{y}{2} - \frac{1}{18}y^2 & 0 \leqslant y \leqslant 3 \\ 1 & y > 3 \end{cases}$

d Median = $\frac{9 - 3\sqrt{5}}{2} = 1.15$ (3 s.f.)

6 a

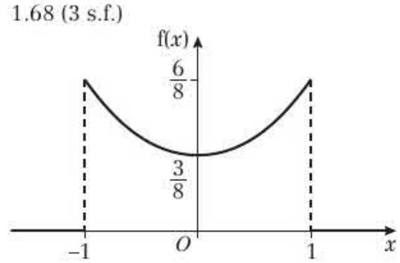


b 2

c $F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{16}x^4 & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

d 1.68 (3 s.f.)

7 a

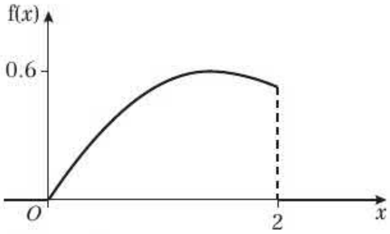


b Bimodal −1 and 1

c Median = 0

d $F(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{8}x^3 + \frac{3}{8}x + \frac{1}{2} & -1 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$

8 a



b Mode = 1.5

c $F(x) = \begin{cases} 0 & x < 0 \\ \frac{9}{20}x^2 - \frac{1}{10}x^3 & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

d F(1.23) = 0.495 and F(1.24) = 0.501. Since 0.5 lies between 0.495 and 0.501 the median lies between 1.23 and 1.24.

9 a $f(x) = \begin{cases} \frac{1}{4}x & 1 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$

b Mode = 3

c $\sqrt{5}$

d $k = 1.9$

10 a $f(x) = \begin{cases} 12x^2(1 - x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$

b mode = $\frac{2}{3}$

c 0.2853

11 a $F(w) = \begin{cases} 0 & w < 0 \\ \frac{w^4}{5^5}(25 - 4w) & 0 \leqslant w \leqslant 5 \\ 1 & w > 5 \end{cases}$

b F(3.4) = 0.487…, F(3.5) = 0.528…, so median lies between 3.4 and 3.5.

c The maximum of f(w) is at $w = \frac{15}{4}$, so the mode is $\frac{15}{4}$

12 a 1.365 (3 d.p.)

b $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{4} & 0 \leqslant x < 1 \\ \frac{x^4}{20} + \frac{1}{5} & 1 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

c Median = 1.565 (3 d.p.) IQR = 0.821 (3 d.p.)

d 1.414

13 a Mode = 2

b $F(x) = \begin{cases} 0 & x < 2 \\ \dfrac{\ln\left(\frac{x}{2}\right)}{\ln 5} & 2 \leqslant x \leqslant 10 \\ 1 & x > 10 \end{cases}$

c $2\sqrt{5}$

d $Q_1 = 2.991$, $Q_3 = 6.687$, IQR = 3.697

14 a 277 hours

b Q_1 = 115 hours, Q_3 = 555 hours, IQR = 439 hours

15 a $k = \pi$

b $F(x) = \begin{cases} 0 & x < 0 \\ \tan(\pi x) & 0 \leqslant x \leqslant 0.25 \\ 1 & x > 0.25 \end{cases}$

c 0.1476

16 a $k = \frac{5}{\ln 6}$ **b** 3.066 (3 d.p.) **c** 0.349 (3 d.p.)

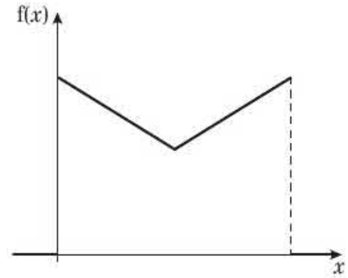
d $F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{\ln 6}\left(\ln\left(\frac{3x}{10 - 2x}\right)\right) & 2 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$

e 3.101 (3 d.p.) **f** 4

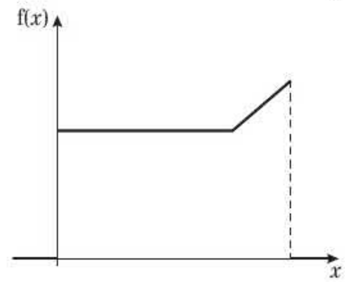
g 2.31

Challenge

1 a e.g.



b e.g.



2 e.g. $f(x) = \begin{cases} x & 0 \leqslant x \leqslant 1 \\ \frac{1}{2} & 1 < x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

Chapter review 4

1 a $\frac{10}{9}, \frac{16}{3}$ **b** $\frac{26}{81}, \frac{26}{9}$ **c** $\frac{5}{12}$

d $\frac{128}{243}$ **e** 0.5

2 a $\frac{1}{3}$ **b** $\frac{1}{18}$ **c** $\frac{5}{3}, \frac{2}{9}$

d $F(x) = \begin{cases} 0 & x < 0 \\ 2x - x^2 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$

e Median = 0.293 as 1.71 is not in the range

3 a F(2) = 1; $F(y) = k(y^2 - y)$

$k(4 - 2) = 1 \Rightarrow k = \frac{1}{2}$

b 0.375

c Median = 1.62 as −0.618 is not in the range

d $f(y) = \begin{cases} y - \frac{1}{2} & 1 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

4 a 0.648

b Median = 2.55 as −2.55 is not in the range

c $f(x) = \begin{cases} \frac{2x}{5} & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$

d $\frac{38}{15}$ **e** Mode = 3

5 a $\int_0^2 kx^2\,dx = 1$; $\left[\frac{kx^3}{3}\right]_0^2 = 1$

$\frac{8k}{3} = 1 \Rightarrow k = \frac{3}{8}$

b 1.5

c $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^3}{8} & 0 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

d 1.59 (3 s.f.) **e** Mode = 2

6 a $\int_1^3 k(y^2 + 2y + 2)\,dx = 1$

$\left[k\left(\frac{y^3}{3} + y^2 + 2y\right)\right]_1^3 = 1$

$k\left(\frac{3^3}{3} + 3^2 + 6\right) - k\left(\frac{1}{3} + 1 + 2\right) = 1$

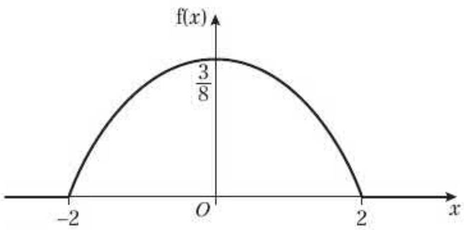
$\frac{62}{3}k = 1$

$k = \frac{3}{62}$

b $F(y) = \begin{cases} 0 & y < 1 \\ \frac{y^3}{62} + \frac{3y^2}{62} + \frac{3y}{31} - \frac{5}{31} & 1 \leqslant y \leqslant 3 \\ 1 & y > 3 \end{cases}$

c $\frac{11}{31}$

7 a



b Mode = 0

c $F(x) = \begin{cases} 0 & x < -2 \\ \frac{12x}{32} - \frac{x^3}{32} + \frac{1}{2} & -2 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

d $\frac{35}{128}$ or 0.273

8 a $\frac{26}{21}$

b $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{3} & 0 \leqslant x < 1 \\ \frac{2x^3}{21} + \frac{5}{21} & 1 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

c i 1.401 **ii** 0.45

9 $F(1) = 0 \Rightarrow 0.05a - b = 0$

$F(2) = 1 \Rightarrow 0.05a^2 - b = 1$

$0.05(a^2 - a) = 1 \Rightarrow a^2 - a - 20 = 0$

$(a + 4)(a - 5) = 0$

Given that a is positive, $a = 5$ and $b = 0.05a = \frac{1}{4}$

10 $F'(x) < 0$ for $8 < x < 10$

11 a $\int_1^3 kx - k\,dx = 1$

$\left[\frac{kx^2}{2} - kx\right]_1^3 = 1$

$\left(\frac{9k}{2} - 3k\right) - \left(\frac{k}{2} - k\right) = 1$

$2k = 1 \Rightarrow k = \frac{1}{2}$

b $\frac{7}{3}$

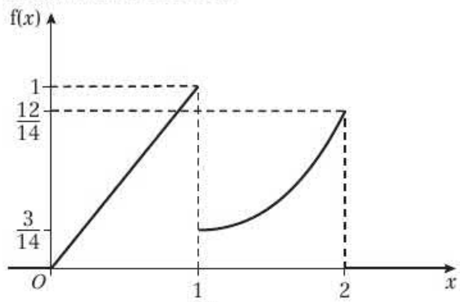
c $F(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2}{4} - \frac{x}{2} + \frac{1}{4} & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$

d $F(2.4) = \frac{2.4^2}{4} - \frac{2.4}{2} + \frac{1}{4} = 0.49$

$F(2.5) = \frac{2.5^2}{4} - \frac{2.5}{2} + \frac{1}{4} = 0.5625$

Since 0.5 lies in between the median is between 2.4 and 2.5

12 a

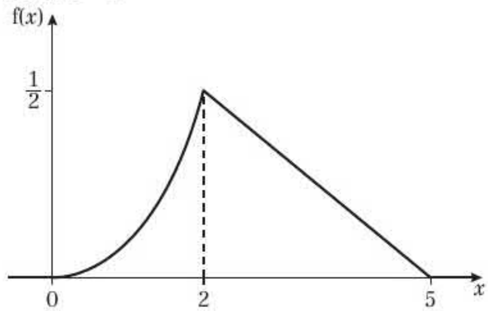


b Mode = 1 **c** $\frac{191}{84}$ **d** 1.14

e $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{2} & 0 < x < 1 \\ \frac{x^3}{14} + \frac{3}{7} & 1 \leqslant x \leqslant 2 \\ 1 & x > 2 \end{cases}$

f Median = 1

13 a



b Mode = 2

c Using the sketch, $P(X > 2)$ = area of triangle $= \frac{1}{2} \times 3 \times \frac{1}{2} = 0.75$

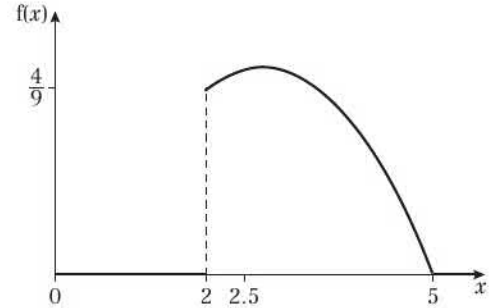
d $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^4}{64} & 0 < x < 1 \\ \frac{10x - x^2 - 13}{12} & 2 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$

e $5 - \sqrt{6} = 2.55$ (s.f.)

14 a $f(x) = \begin{cases} \frac{1}{81}(-6x^2 + 30x) & 2 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$

b Mode = 2.5

c



d $\frac{19}{6}$

e $F(\mu) = F\left(\frac{19}{6}\right) = \frac{1}{81}\left[-2\left(\frac{19}{6}\right)^3 + 15\left(\frac{19}{6}\right)^2 - 44\right]$

$= 0.5297 \text{ (4 d.p.)} > 0.5$

f $F(2.5) = 0.2284$ so as $0.2284 < 0.5 < 0.5297$, for this distribution mode < median < mean

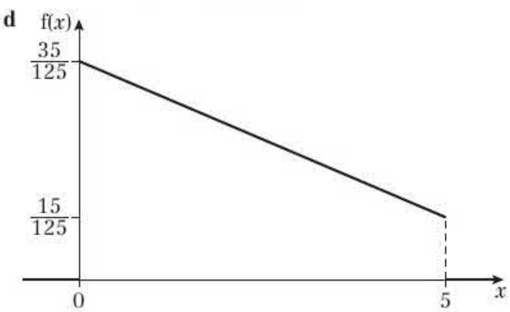
15 a $F(5) = 1 \Rightarrow k(35 \times 5 - 2 \times 5^2) = 1 \Rightarrow 125k = 1$

$\Rightarrow k = \frac{1}{125}$

b 2.02 (3 s.f.)

c $f(x) = \begin{cases} \frac{35 - 4x}{125} & 0 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$

d



e Mode = 0

f $\frac{13}{6}$

g 0.180

16 $a = \frac{3}{16}, b = \frac{5}{16}$

17 a $\int_{-1}^{0} k(x+1)^3 \, dx = 1 \Rightarrow \left[\frac{k(x+1)^4}{4}\right]_{-1}^{0} = 1$

$\Rightarrow \frac{k}{4} = 1 \Rightarrow k = 4$

b −0.2

c $F(x) = \begin{cases} 0 & x < -1 \\ (x+1)^4 & -1 \leqslant x \leqslant 0 \\ 1 & x > 0 \end{cases}$

d −0.159 (3 s.f.)

18 a $F(t) = \begin{cases} 0 & t < 0 \\ 1 - \frac{(6-t)^3}{216} & 0 \leqslant t \leqslant 6 \\ 1 & t > 6 \end{cases}$

b 1.24 hours (3 s.f.) **c** 1.5 hours

19 $F(x) = \begin{cases} 0 & x < 1 \\ \frac{\ln(2x-1)}{\ln 5} & 1 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$

20 a $k = \pi$

b 0.5947 = 59.47%

21 a $k = \frac{2}{3}$

b $\frac{1}{3} + \frac{2}{3}\ln 2 = 0.795$ (3 s.f.)

c 0.256 (3 s.f.)

Challenge

a $E(X) = \int_0^\infty x f(x) \, dx = \int_0^\infty x(\lambda e^{-\lambda x}) \, dx$

$= [x(\lambda e^{-\lambda x})]_0^\infty - \int_0^\infty -e^{-\lambda x} \, dx$

$= 0 + \int_0^\infty e^{-\lambda x} \, dx = -\frac{1}{\lambda}(0 - 1) = \frac{1}{\lambda}$

$E(X^2) = \int_0^\infty x^2 f(x) \, dx = \int_0^\infty x^2(\lambda e^{-\lambda x}) \, dx$

$= [x^2(\lambda e^{-\lambda x})]_0^\infty - \int_0^\infty -2x e^{-\lambda x} \, dx$

$= 0 + 2\int_0^\infty x e^{-\lambda x} \, dx = \frac{2}{\lambda^2}$

$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$

b $P(X > a) = 1 - P(X < a) = 1 - \int_0^a \lambda e^{-\lambda x} \, dx$

$= 1 - [-e^{-\lambda x}]_0^a = 1 - (-e^{-\lambda a} + 1)0 = e^{-\lambda a}$

Similarly, $P(X > b) = e^{-\lambda b}$ and $P(X > a + b) = e^{-\lambda(a+b)}$

$= e^{-\lambda a} \times e^{-\lambda b}$

$P(X > a + b \mid X > a) = \frac{P(X > a + b)}{P(X > a)}$

$= \frac{e^{-\lambda a} \times e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda b} = P(X > b)$

Review exercise 1

1 a $E(X) = 4$, $\text{Var}(X) = 3.92$

b n is large and p is small. Also $E(X) \approx \text{Var}(X)$

c 0.785

2 a $X \sim B(20, 0.2)$

b 0.196

c \$196.20

3 a 0.1011 **b** 0.4477 **c** 0.9289 **d** 0.3393

4 a $X \sim \text{Po}(1.5)$ **b** 0.251 (3 s.f.)

c 0.469 (3 s.f.) **d** 0.185 (3 s.f.)

5 a Events occur at a constant rate.
Events occur independently or randomly.
Events occur singly.

b i 0.134 (3 s.f.) **ii** 0.715 (3 s.f.)

c 0.149 (3 s.f.)

6 a 0.0816 **b** 0.1931 **c** 0.5673

7 a 1.45; 1.4075

b Mean ≈ Variance

c $P(X > 2) = 0.4253$

8 a 0.1528

b 6

9 a 0.00248

b 0.671

10 a If $X \sim B(n, p)$ and
- n is large
- p is small

then X can be approximated by $\text{Po}(np)$.

b 0.0001

c 0.00098

d mean $= np = 10$
variance $= np(1 - p) = 9.9$

e 0.870 (3 s.f.)

11 a 0.226 (3 s.f.)

b $\lambda = 3$. If $X \sim B(n, p)$, n is large, and p is small, then X can be approximated by $\text{Po}(np)$.

12 a $X \sim B(200, 0.015)$

b 0.1693

c If $X \sim B(n, p)$, n is large, and p is small, then X can be approximated by $Po(np)$.
$\lambda = 3$

d $P(X = 4) = 0.1680$
% error = 0.77

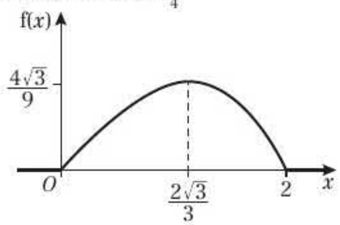
13 0.541

14 0.7054

15 0.3539

16 a $\int_0^2 k(4x - x^3)\,dx = 1 \Rightarrow \left[2kx^2 - \frac{k}{4}x^4\right]_0^2 = 1$
$\Rightarrow 4k = 1 \Rightarrow k = \frac{1}{4}$

b



c $E(X) = \frac{16}{15}$ (or 1.07 to 3 s.f.)

d Mode $= \frac{2\sqrt{3}}{3}$ (or 1.15 to 3 s.f.)

e Median = 1.08 (3 s.f.)

17 a $\int_2^3 kx(x-2)dx = 1 \quad \Rightarrow \quad k\left[\frac{1}{3}x^3 - x^2\right]_2^3 = 1$
$k\left((9-9) - \left(\frac{8}{3} - 4\right)\right) = 1 \quad \Rightarrow \quad k = \frac{3}{4}$

b $\frac{67}{1280}$

c $F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{4}(x^3 - 3x^2 + 4) & 2 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$

d $F(2.70) = 0.453$ and $F(2.75) = 0.527$
$0.453 < 0.5 < 0.527$ so the median lies between 2.70 and 2.75

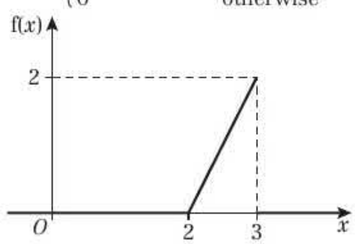
18 a $F'(y) < 0$ for $1.625 < y < 2$ so his model cannot be a cumulative distribution function.

b $k(2^4 + 2^2 - 2) - k(1 + 1 - 2) = 1$
$\Rightarrow k(16 + 4 - 2) = 1 \Rightarrow 18k = 1 \Rightarrow k = \frac{1}{18}$

c $\frac{203}{288}$ or 0.705 (3 s.f.)

d $f(y) = \begin{cases} \frac{1}{9}(2y^3 + y) & 1 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

19 a

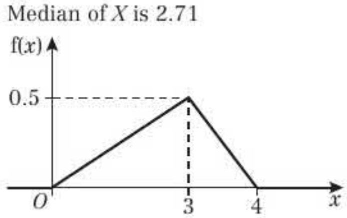


b Mode of X is 3.

c $\frac{1}{18}$

d Median of X is 2.71

20 a



b Mode of X is 3.

c $F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{12}x^2 & 0 \leqslant x < 3 \\ 2x - \frac{1}{4}x^2 - 3 & 3 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$

d Median $= \sqrt{6} = 2.45$ (3 s.f.)

e 2.272 (3 d.p.)

21 a 0.847

b $F(0.59) = 0.491$ and $F(0.60) = 0.504$
$0.491 < 0.5 < 0.504$ so the median lies between 0.59 and 0.60

c $f(x) = \begin{cases} 4x - 3x^2 & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$

d $E(X) = \frac{7}{12}$ or 0.583 (3 s.f.)

e Mode $= \frac{2}{3}$ or 0.667 (3 s.f.)

22 a $\int_0^2 k\,dx + \int_2^4 \frac{k}{x}dx = 1$
$\Rightarrow [kx]_0^2 + [k \ln x]_2^4 = 1$
$\Rightarrow k(2 + \ln 2) = 1$
$\Rightarrow k = \frac{1}{2 + \ln 2}$

b $\frac{4}{2 + \ln 2}$ (= 1.485...)

23 a $\frac{33}{49}$ **b** $\frac{11}{35}$ **c** $\frac{23}{3}$

24 a $\int_1^5 \frac{1}{148}(x^2 + 2)dx + \int_5^7 k(3x - 5)dx = 1$
$\left[\frac{1}{148}\left(\frac{x^3}{3} + 2x\right)\right]_1^5 + k\left[\frac{3x^2}{2} - 5x\right]_5^7 = 1$
$\frac{1}{3} + 26k = 1$
$26k = \frac{2}{3}$
$k = \frac{1}{39}$

b $F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{444}(x^3 + 6x - 7) & 1 \leqslant x \leqslant 5 \\ \frac{1}{78}(3x^2 - 10x + 1) & 5 \leqslant x \leqslant 7 \\ 1 & x > 7 \end{cases}$

c 5.64 **d** 6.49

Challenge

a $\int_0^\infty ke^{-x}dx = k[-e^{-x}]_0^\infty = k(=1) \Rightarrow k = 1$

b $F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geqslant 0 \end{cases}$

c $e^{-1} - e^{-4}$ or $\frac{e^3 - 1}{e^4}$

CHAPTER 5

Prior knowledge check

1 a $k = \frac{1}{20}$ **b** $\frac{7}{10}$ **c** 5

2 a 4 **b** 3 **c** 48

3 a $a = 6$

Exercise 5A

1 a 0.4 **b** 0.6

2 a $k = 12.6$ **b** 0.39

3 a $k = \frac{1}{8}$ **b** 0.6875 **c** $p = 4$
d $\frac{1}{6}$ **e** $\frac{3}{5}$ **f** $\frac{1}{2}$

4 $a = 3, b = 11$

5 **a** $Y \sim U[9, 21]$ **b** $\frac{2}{3}$

6 **a** Continuous uniform distribution
b $E(Y) = 6$ **c** $\frac{2}{5}$ **d** $\frac{3}{4}$

7 **a** 1 **b** $\frac{16}{3}$ **c** $\frac{19}{3}$
d
$$F(x) = \begin{cases} 0 & x < -3 \\ \dfrac{x+3}{8} & -3 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$

8 **a** $E(X) = 3, \text{Var}(X) = \frac{4}{3}$ **b** $E(X) = 2, \text{Var}(X) = \frac{16}{3}$

9 **a** 4.5 **b** $\frac{1}{3}$ **c** $20\frac{7}{12} = 20.6$
d
$$F(x) = \begin{cases} 0 & x < 3.5 \\ \dfrac{x}{2} - 1.75 & 3.5 \leqslant x \leqslant 5.5 \\ 1 & x > 5.5 \end{cases}$$

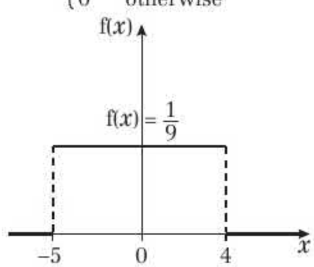
10 $a = -1$ and $b = 3$

11 $E(X) = \dfrac{5 + (-1)}{2} = 2$
$\text{Var}(X) = \dfrac{(5 - (-1))^2}{12} = 3$
$E(Y) = 4E(X) - 6$
$= 8 - 6 = 2$
$\text{Var}(Y) = 16\,\text{Var}(X)$
$= 48$

12 **a** $\alpha = 3, \beta = 7$ **b** 0.55

13 **a** $f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha \leqslant x \leqslant \beta \\ 0 & \text{otherwise} \end{cases}$
b $\alpha = -3, \beta = 8$

14 **a** $f(x) = \begin{cases} \frac{1}{9} & -5 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$
b



c 7 **d** $\frac{4}{45}$

15 **a** $\frac{3}{7}$ **b** $f(x) = \begin{cases} \frac{1}{7} & -3 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$
c Uniform **d** Mean = 0.5, Variance = $\frac{49}{12}$

16 **a** 1.5 **b** $\frac{25}{12}$ **c** $\frac{13}{3}$
d 0.48 **e** 0.2153

17 **a** $\alpha = 1.5, \beta = 13.5$ **b** **i** $c = 5.5$ **ii** $\frac{7}{24}$

Challenge
a $\frac{4}{7}$ **b** $\frac{2}{5}$

Exercise 5B

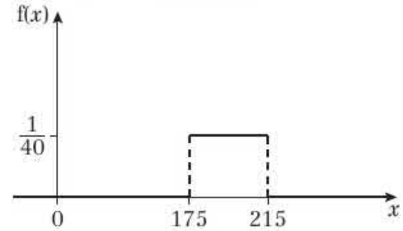
1 $E(Y) = E(X^2) = 25\frac{1}{12}$

2 **a** $f(x) = \begin{cases} \frac{1}{6} & 5 \leqslant x \leqslant 11 \\ 0 & \text{otherwise} \end{cases}$
b 0.5 **c** $67\pi\,\text{cm}^2$

3 **a** 0.2 **b** 0.5 **c** $\frac{1}{12}$

4 **a** $\frac{3}{8}$ **b** $\frac{27}{512}$ **c** $\frac{3}{5}$

5 **a** $f(x) = \begin{cases} \frac{1}{40} & 175 \leqslant x \leqslant 215 \\ 0 & \text{otherwise} \end{cases}$



b **i** 0.3 **ii** 0
c 20 **d** 189 **e** 0.1323 (4 d.p.)

6 **a** $\frac{7}{60}$ **b** $\frac{1}{3}$ **c** 0.2276 (4 d.p.)

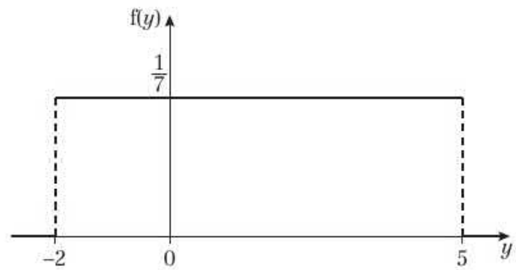
7 **a** $\frac{1}{4}$ **b** 0.2142 (4 d.p.)

8 **a** 0.6 **b** 0.3222 (4 d.p.)

9 $\frac{200}{3}$

Chapter review 5

1 **a**



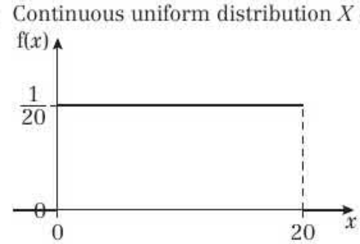
b 1.5 **c** $4\frac{1}{12}$
d
$$F(x) = \begin{cases} 0 & x < -2 \\ \dfrac{x+2}{7} & -2 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$
e $\frac{3}{14}$ **f** 0 **g** $\frac{1}{2}$ **h** $\frac{2}{5}$

2 **a** $k = 1$ **b** 0.2 **c** -1.5 **d** $\frac{25}{12}$
e
$$F(x) = \begin{cases} 0 & x < -4 \\ \dfrac{x+4}{5} & -4 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

3 **a** $a = -1, b = 5$
b 0.533 (3 s.f.)

4 **a** Continuous uniform distribution $Y \sim U[0, 10]$
b $\frac{3}{10}$ **c** $\frac{3}{5}$

5 **a** Continuous uniform distribution $X \sim U[0, 20]$



b $E(X) = 10$ $\text{Var}(X) = \frac{100}{3}$ **c** 0.6

6 **a** $X \sim U[-0.5, 0.5]$ **b** 0.4 **c** $\frac{1}{12}$

7 a $f(x) = \begin{cases} \frac{1}{13} & -3 \leqslant x \leqslant 10 \\ 0 & \text{otherwise} \end{cases}$

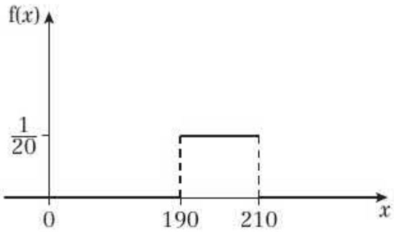
b 3.5 minutes

c $F(x) = \begin{cases} 0 & x < -3 \\ \frac{x+3}{13} & -3 \leqslant x \leqslant 10 \\ 1 & x > 10 \end{cases}$

d $\frac{5}{13}$

8 a $X \sim U[-0.5, 0.5]$ **b** 0.4 **c** 0.064

9 a $f(x) = \begin{cases} \frac{1}{20} & 190 \leqslant x \leqslant 210 \\ 0 & \text{otherwise} \end{cases}$



b i $\frac{2}{5}$ **ii** 0 **c** 10 **d** $\frac{2}{3}$

10 a Continuous uniform distribution

b Normal distribution

11 a $f(x) = \begin{cases} \frac{1}{4b} & b \leqslant x \leqslant 5b \\ 0 & \text{otherwise} \end{cases}$

b $3b$

c $E(X^2) = \int_b^{5b} \frac{x^2}{4b}\,dx = \frac{1}{4b}\left[\frac{x^3}{3}\right]_b^{5b} = \frac{1}{4b}\left[\frac{125b^3 - b^3}{3}\right]$

$= \frac{124b^2}{12} = \frac{31b^2}{3}$

d $\frac{5}{12}$

e 0.246 (3 s.f.)

Challenge

1 a $\theta \sim U[0, 2\pi]$

b π

c Spin the spinner 100 times and measure X each time. Take the mean of these observations and divide $2r$ by this value.

CHAPTER 6

Prior knowledge check

1 Mean = 18.25, Median =18.5, Mode = 20; Range = 8

2 a HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

b

h	0	1	2	3
$P(H = h)$	0.125	0.375	0.375	0.125

Exercise 6A

1 a A census observes or measures every member of a population.

b Advantage: will give a completely accurate result. Disadvantages: time-consuming, expensive, not practical

2 a The testing process will destroy the rope, so a census would destroy *all* the ropes.

b 250 kg is the median load at which the ropes in the sample break. This means that half of the ropes will break at a load less than 250 kg.

c Test a larger number of ropes.

3 a Any one from:
It would be expensive.
It would be time consuming.
It would be difficult.

b A list of residents. **c** A resident.

4 a The testing process will destroy the microswitches, so a census would destroy *all* the switches.

b The mean is less than the stated average but one of the switches lasted a significantly lower number of operations which suggests the median might be a better average to take – not affected by outliers. The data supports the company claim.

c Test a larger number of microswitches.

5 a All the mechanics in the garage.

b Everyone's views will be known.

Exercise 6B

1 This mean is from the value of a sample so it is a statistic.

2 **i** and **ii** are statistics.

iii is not a statistic because it uses μ.

3 a All the hairdressers who work for the chain of hairdressing shops.
The proportion p of the staff who are happy to wear an apron.

b This is a binomial distribution since we are only interested in two options – whether the hairdressers are happy or not.

4 a Po(3) **b** 0.1991

5 a $\mu = 32.5$; $\sigma = 318.75$

b (50, 50) (50, 20) (20, 50) (50, 10) (10, 50) (20, 20) (20, 10) (10, 20) (10, 10)

c

x	50	35	30	20	15	10
$P(X = x)$	0.25	0.25	0.25	0.0625	0.125	0.0625

6 a Mean = 19.4 Variance = 16.04

b (16, 16) (16, 20) (20, 16) (16, 30) (30, 16) (30, 30) (30, 20) (20, 30) (20, 20)

c

x	16	18	20	23	25	30
$P(X = x)$	0.16	0.4	0.25	0.08	0.01	0.01

7 a Mean = 2.6; Variance = 0.24

b (3, 3, 3) (3, 3, 2) (3, 2, 3) (2, 3, 3) (3, 2, 2) (2, 3, 2) (2, 2, 3) (2, 2, 2)

c

x	3	$2\frac{2}{3}$	$2\frac{1}{3}$	2
$P(X = x)$	0.216	0.432	0.288	0.064

d

M	3	2
$P(M = m)$	0.648	0.352

e

N	3	2
$P(N = n)$	0.648	0.352

Chapter review 6

1 a A list of all the patients on the surgery books.

b A patient.

2 a Any two from:
It would take too long.
It would cost too much.
It could be difficult to get hold of all members.

b A list of all members of the gym.

c A member of the gym.

3 a A sampling frame has to be some sort of list – it may not be possible to list a population.

b A sample is usually easier to do, quicker to do and not as costly as a census.

4 a A statistic is a quantity calculated solely from the observations of a sample.
b i is a statistic
ii is not a statistic as it depends on the value μ.

5 a The light bulbs would all be destroyed.
b A light bulb.

6 a Any two from:
It is quicker to do.
It is cheaper to do.
It is easier to do.
b It can be biased OR it is subject to natural variations.
c A numbered list of all 400 call centre operators.
d A call-centre operator.
e Yes, because he is using only the values from a sample. There are no parameters.

7 Any two from:
It gives everyone's views.
It is unbiased.
To take a sample when the population is only 10 would be silly.

8 **a** and **b** are statistics
c and **d** are not statistics since they involve a population parameter.

9 a Mean = 9 1/6 Variance = 28.47
b (5, 5) (10, 10) (20, 20) (5, 10) (10, 5) (5, 20) (20, 5) (10, 20) (20, 10)
c

y	5	7.5	10	12.5	15	20
$P(Y = y)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{36}$

10 a (6, 6, 6) (6, 6, 10) (6, 10, 6) (10, 6, 6) (6, 10, 10) (10, 6, 10) (10, 10, 6) (10, 10, 10)
b

M	3	2
$P(M = m)$	0.648	0.352

c

N	3	2
$P(N = n)$	0.648	0.352

Challenge

a $\text{E}(\overline{X}) = \text{E}\left(\frac{X_1 + X_2 + \cdots X_n}{n}\right)$
$= \frac{1}{n}\text{E}(X_1 + \cdots + X_n)$
$= \frac{1}{n}-(\text{E}(X_1) + \cdots + \text{E}(X_n))$
$= \frac{1}{n}(\mu + \cdots + \mu) = \frac{1}{n}(n\mu) = \mu$

b $\text{Var}(\overline{X}) = \text{Var}\left(\frac{X_1 + X_2 + X_3}{3}\right)$
$= \frac{1}{9}\text{Var}(X_1 + X_2 + X_3)$
$= \frac{1}{9}(\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3))$
$= \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2) = \frac{\sigma^2}{3}$

CHAPTER 7

Prior knowledge check

1 a 0.075 **b** 0.117
c 0.0036 **d** 0.00000504

2 a $X \sim B(8, \frac{1}{6})$ **b i** 0.260 **ii** 0.0307

Exercise 7A

1 a A hypothesis is a statement made about the value of a population parameter. A hypothesis test uses a sample or an experiment to determine whether or not to reject the hypothesis.
b The null hypothesis (H_0) is what we assume to be correct and the alternative hypothesis (H_1) tells us about the parameter if our assumption is shown to be wrong.
c A test statistic is used to test the hypothesis. It could be the result of the experiment or statistics calculated from a sample.

2 a One-tailed test
b Two-tailed test
c One-tailed test

3 a The test statistic is N – the number of sixes.
b H_0: $p = \frac{1}{6}$ **c** H_1: $p > \frac{1}{6}$

4 a Isabelle is describing what her experiment wants to test rather than the test statistic. The test statistic is the proportion of times the coin lands on heads.
b H_0: $p = \frac{1}{2}$ **c** H_1: $p \neq \frac{1}{2}$

5 a A suitable test statistic is p – the proportion of faulty articles in a batch.
b H_0: $p = 0.1$, H_1: $p < 0.1$
c If the probability of the proportion being 0.08 or less is 5% or less, the null hypothesis is rejected.

6 a A suitable test statistic is p – the proportion of people who vote for the group.
b H_0: $p = 0.55$, H_1: $p < 0.55$
c If the probability of the proportion being $\frac{7}{20}$ is 2% or more, the null hypothesis is accepted

Exercise 7B

1 a The critical value is the first value to fall inside of the critical region.
b A critical region is a region of the probability distribution which, if the test statistic falls within it, would cause you to reject the null hypothesis.
c The acceptance region is the area in which we accept the null hypothesis.

2 The critical value is $x = 5$ and the critical region is $X \geqslant 5$ since $P(X \geqslant 5) = 0.0328 < 0.05$

3 The critical value is $x = 0$ and the critical region is $X = 0$.

4 a The critical region is $X \geqslant 13$ and $X \leqslant 3$.
b $0.037 = 3.7\%$

5 The critical value is $x = 0$. The critical region is $X = 0$.

6 a The critical region is $X = 0$ and $7 \leqslant X \leqslant 10$.
b 0.085

7 a The number of times the sample fails.
b H_0: $p = 0.3$, H_1: $p < 0.3$
c The critical value is $x = 10$ and the critical region is $X \geqslant 10$
d 4.8%

8 a The number of seedlings that survive.
H_0: $p = \frac{1}{3}$, H_1: $p > \frac{1}{3}$
b The critical value is $x = 17$ and the critical region is $X \geqslant 17$
c 5.84%

9 a H_0: $p = 0.2$, H_1: $p \neq 0.2$
b The critical region is $X \leqslant 1$ and $X \geqslant 10$
c 4.47%

Challenge

a The critical region is $X \leqslant 29$ and $X \geqslant 41$

b Chance of one observation falling within the critical region = 8.8%
Chance of two observations falling within the critical region = 0.77%

Exercise 7C

1 $0.0781 > 0.05$
There is insufficient evidence to reject H_0.

2 $0.0464 < 0.05$
There is sufficient evidence to reject H_0 so $p < 0.04$

3 $0.0480 < 0.05$
There is sufficient evidence to reject H_0 so $p > 0.30$

4 $0.0049 < 0.01$
There is sufficient evidence to reject H_0 so $p < 0.45$

5 $0.0526 > 0.05$
There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.28$

6 $0.0020 < 0.05$
There is sufficient evidence to reject H_0 so $p > 0.32$

7 $0.3813 > 0.05$
There is insufficient evidence to reject H_0 (not significant).
There is no evidence that the probability is less than $\frac{1}{6}$.
There is no evidence that the dice is biased.

8 **a** Distribution $B(n, 0.68)$
Fixed number of trials.
Outcomes of trials are independent.
There are two outcomes success and failure.
The probability of success is constant.

b $P(X \leqslant 3) = 0.0155 < 0.05$. There is sufficient evidence to reject the null hypothesis so $p < 0.68$
The treatment is not as effective as claimed.

9 **a** Critical region is $X \geqslant 13$

b 14 lies in the critical region, so we can reject the null hypothesis. There is evidence that the new technique has improved the number of plants that grow.

10 **a** The number of people who support the candidate.
H_0: $p = 0.35$, H_1: $p > 0.35$

b Critical region is $X \geqslant 24$

c 28 lies in the critical region, so we can reject the null hypothesis. There is evidence that the candidate's level of popularity has increased.

Exercise 7D

1 $P(X \leqslant 10) = (0.0494 > 0.025$ (two-tailed)
There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.5$

2 $P(X \geqslant 10) = 0.189 > 0.05$ (two-tailed)
There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.3$

3 $(X \geqslant 9) = 0.244 > 0.025$ (two-tailed)
There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.75$

4 $P(X \leqslant 1) = 0.00000034 < 0.005$ (two-tailed)
$X = 1$ lies within the critical region, so we can reject the null hypothesis.

5 $P(X \geqslant 4) = 0.0178 > 0.01$ (two-tailed)
There is insufficient evidence to reject H_0 so there is no reason to doubt $p = 0.02$

6 $P(X \leqslant 6) = 0.0577 > 0.025$ (two-tailed)
$X = 6$ does not lie in the critical region, so there is no reason to think that the coin is biased.

7 **a** Critical region $X = 0$ and $X \geqslant 8$

b 4.36%

c H_0: $p = 0.2$, H_1: $p \neq 0.2$
$X = 8$ is in the critical region. There is enough evidence to reject H_0. The hospital's proportion of complications differs from the national figure.

8 Test statistic: the number of cracked bowls.
H_0: $p = 0.1$, H_1: $p \neq 0.1$
$P(X \leqslant 1) = 0.3917 = 39.17\%$
$39.17\% > 5\%$ (two-tailed) so there is not enough evidence to reject H_0. The proportion of cracked bowls has not changed.

9 Test statistic: the number of carrots longer than 7 cm
H_0: $p = 0.25$, H_1: $p \neq 0.25$
$P(X \geqslant 13) = 1 - P(X \leqslant 12) = 0.0216 = 2.16\%$
$2.16\% < 2.5\%$ (two-tailed) so there is enough evidence to reject the null hypothesis. The probability of a carrot being longer than 7 cm has increased.

10 Test statistic: the number of patients correctly identified.
H_0: $p = 0.96$, H_1: $p \neq 0.96$
$P(X \leqslant 63) = 0.0000417 < 0.05$ (two-tailed) so there is enough evidence to reject the null hypothesis. The new test does not have the same probability of success as the old test.

Exercise 7E

1 Critical region is $X \geqslant 7$

2 Critical region is $X \geqslant 9$

3 Critical region is $X \leqslant 2$

4 $0.2378 > 0.025$; do not reject H_0

5 $0.184 > 0.05$; do not reject H_0

6 $0.2851 > 0.05$; do not reject H_0

7 $0.0511 < 0.1$; reject H_0

8 **a** H_0: $\lambda = 4$, H_1: $\lambda > 4$

b Critical region $X \geqslant 9$

c 8 is not in the critical region. The scientist concluded that there was not enough evidence to suggest an increase in the number of storms.

9 There is evidence that the weekly sales has decreased.

10 Reject H_0
There is evidence that the percentage of workers who are absent for at least 1 day per month is less than 20%

Chapter review 7

1 H_0: $p = 0.2$, H_1: $p > 0.2$, $P(X \geqslant 3) = 0.3222 > 0.05$
There is insufficient evidence to reject H_0.
There is no evidence that the trains are late more often.

2 H_0: $p = 0.5$, H_1: $p > 0.5$, $P(X \geqslant 4) = 0.1875 > 0.05$
There is insufficient evidence to reject H_0.
There is insufficient evidence that the manufacturer's claim is true.

3 **a** Fixed number; independent trials; two outcomes (pass or fail); p constant for each car.

b 0.16807

c $0.3828 > 0.05$
There is insufficient evidence to reject H_0.
There is no evidence that the garage fails fewer than the national average.

4 **a** Critical region $X \leqslant 1$ and $X \geqslant 10$

b 0.0583

c H_0: $p = 0.1$, H_1: $p > 0.1$, $P(X \geqslant 4) = 0.133 > 0.1$
Accept H_0. There is no evidence that the proportion of faulty articles has increased.

5 $H_0: p = 0.5$, $H_1: p \neq 0.5$,
$P(X \leqslant 8) = 0.252 > 0.025$ (two-tailed)
There is insufficient evidence to reject H_0.
There is no evidence that the claim is wrong.

6 a Critical region is $X \leqslant 4$ and $X \geqslant 16$
b 0.0493
c There is insufficient evidence to reject H_0. There is no evidence to suggest that the proportion of people buying that certain model of computer differs from 0.2.

7 a i The theory, methods, and practice of testing a hypothesis by comparing it with the null hypothesis.
ii The critical value is the first value to fall inside of the critical region.
iii The acceptance region is the region where we accept the null hypothesis.
b Critical region $X = 0$ and $X \geqslant 8$
c 4.36%
d As 7 does not lie in the critical region, H_0 is not rejected. Therefore, the proportion of times that Johan is late for school has not changed.

8 $P(X \geqslant 21) = 0.021 < 0.05$. Therefore there is sufficient evidence to support Clara's claim that the likelihood of a rain-free day has increased.

9 a Critical region $X \leqslant 5$ and $X \geqslant 16$
b 5.34%
c $X = 4$ is in the critical region so there is enough evidence to reject H_0.

10 a $X \sim B(20, 0.85)$
b 0.1821
c Test statistic is proportion of patients who recover.
$H_0: p = 0.85$, $H_1: p < 0.85$
$P(X \leqslant 20) = 0.00966$
$0.00966 < 0.05$ so there is enough evidence to reject H_0. The percentage of patients who recover after treatment with the new cream is lower than 85%.

11 a A hypothesis test about a population parameter p tests a null hypothesis H_0 which specifies a particular value for p, against an alternative hypothesis H_1 which is that p has increased, decreased or changed. H_1 will indicate whether the test is one- or two-tailed.
b You can count the number of cups of tea that were served in a given time interval
c Critical region is $X \geq 6$
d 0.0166

12 a 0.1339
b There is sufficient evidence to reject H_0. There is evidence that the rate of hiring caravans has increased.

13 $0.0282 < 0.05$, reject H_0
There is evidence that the filter is failing to work properly.

14 $0.0011 < 0.05$, reject H_0
There is evidence that the rate of sales of onion marmalade has increased after the program.

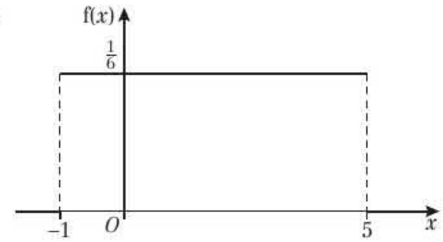
15 $0.0415 < 0.05$, reject H_0
There is evidence that the process is getting worse.

16 Reject H_0
The new variety is better.

Review exercise 2

1 a $a = 7$, $b = 55$
b 0.6

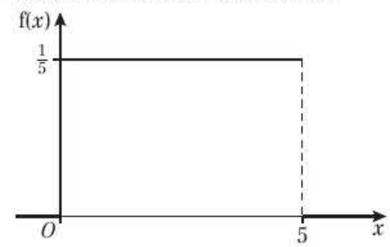
2 a



b $E(X) = 2$
c $Var(X) = 3$
d 0.6

3 a $f(x) = \begin{cases} \frac{1}{4} & 2 \leqslant x \leqslant 6 \\ 0 & \text{otherwise} \end{cases}$
b $E(X) = 4$
c $Var(X) = \frac{4}{3}$
d $F(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{4}(x-2) & 2 \leqslant x \leqslant 6 \\ 1 & x > 6 \end{cases}$
e 0.275

4 a Continuous uniform distribution



b $E(X) = 2.5$; $Var(X) = \frac{25}{12}$
c $\frac{2}{5}$ **d** 0

5 a $f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha \leqslant x \leqslant \beta \\ 0 & \text{otherwise} \end{cases}$
b $\alpha = -2$ $\beta = 6$

6 a 75 cm **b** 43.3 (3 s.f.) **c** $\frac{60}{150} = \frac{2}{5}$

7 a $a = 7$, $b = 17$
b $k = 4$
c $152\frac{1}{3}$
d 0.9

8 a i A named or numbered list of all members of the population.
ii A random variable consisting of any function of the observations and no other quantities.
b i A statistic.
Contains only observations.
ii Not a statistic.
Contains a parameter μ

9 a (R, R) (R, B) (R, Y) (B, R) (B, B) (B, Y) (Y, R) (Y, B) (Y, Y)
b Let X be the number of points gained.

x	1	5
$P(X = x)$	$\frac{11}{18}$	$\frac{7}{18}$

c 46

10

m	1	2	5
$P(M = m)$	0.216	0.352	0.432

11 a i A hypothesis test is where the value of a population parameter (whose assumed value is given in H_0) is tested against what value it takes if H_0 is rejected (this could be an increase, a decrease or a change).

ii A range of values of a test statistic that would lead to the rejection of the null hypothesis

b Critical region $X \leqslant 3, X \geqslant 16$

c 0.0432 or 4.32%

d Insufficient evidence to reject H_0
The rate of incoming calls is less during the school holidays is not supported.

12 There is significant evidence that near the factory the river is polluted with bacteria at the 5% level.

13 a Critical region is $X = 0$ or $X \geqslant 7$

b 0.0607

c There is no evidence to reject the null hypothesis. The probability that a pin chosen at random is not less than 0.15

14 a $0 \leqslant X \leqslant 5$ and $19 \leqslant X \leqslant 40$

b 0.0234

15 a $X \sim B(10, 0.75)$
where X is the random variable 'number of patients who recover when treated'

b 0.146

c H_0: $p = 0.75$ H_1: $p < 0.75$. $0.2142 > 0.05$ so there is insufficient evidence to reject H_0.

d 9

16 a H_0: $p = 0.3$, H_1: $p > 0.3$

b $18 \leqslant X \leqslant 40$

c 3.2%

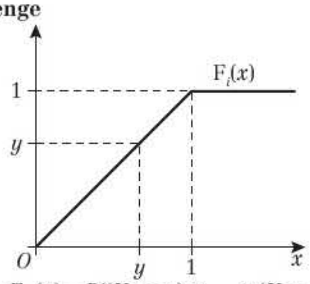
d Reject the null hypothesis. Dhriti's claim is supported.

17 a $X \leqslant 102$

b 115 is not in the critical region so there is no evidence to doubt Theo's claim.

Challenge

1 a



$F_Y(y) = P((X_1 \leqslant y) \cap \ldots \cap (X_n \leqslant y))$

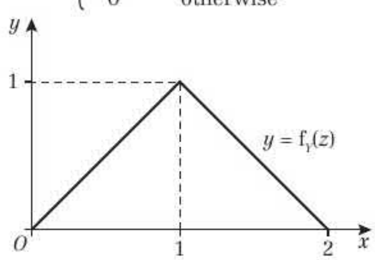
$= F_1(y) \ldots F_n(y) = y^n$

So $f_y(y) = ny^{n-1}$

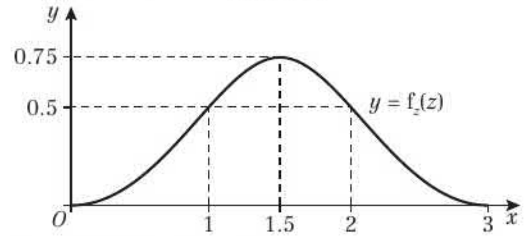
$\Rightarrow E(Y) = \int_0^1 y(ny^{n-1})\,dy = n\int_0^1 y^n\,dy = \frac{n}{n+1}$

b $\sqrt[n]{0.5}$

c $f_Y(z) = \begin{cases} z & 0 \leqslant z \leqslant 1 \\ 2 - z & 1 < z \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$



d $f_z(z) = \begin{cases} \frac{1}{2}z^2 & 0 \leqslant z \leqslant 1 \\ \frac{3}{4} - \left(z - \frac{3}{2}\right)^2 & 1 < z \leqslant 2 \\ \frac{1}{2}(z-3)^2 & 2 < z \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$



2 a $X \leqslant 16$ (probability = 0.1263)

b 0.0160

Exam practice

1 a i 0.172

ii 0.0046

2 a Not a statistic – it is not a function of solely values from the sample. It contains a parameter μ

b A statistic – it is a function of solely values from the sample.

3 a 0.1991

b 0.1606

c 0.00157

4 b $a = -2, b = 16$

c $F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{33}(-x^2 + 16x - 15) & 1 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$

d mode = 1

5 a $f(x) = \begin{cases} \frac{1}{80} & 0 \leqslant x \leqslant 80 \\ 0 & \text{otherwise} \end{cases}$

b 0.125

c 0.3

6 a i 0.1396

ii 0.6866

b $n = 9$

7 Critical value is 15. So reject H_0. There is sufficient evidence to suggest that the journal has underestimated the probability.

8 b

x	0	1	2	3	4	6
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

c Mode = 0

9 a n large, p small

b H_0: $p = 0.004$
H_1: $p > 0.004$
Reject H_0
There is sufficient evidence to suggest that the proportion of faulty lightbulbs has increased.

Online Worked solutions are available in SolutionBank.

INDEX

A

B

C

D

E

G

H

I

Q

R

S

T

U

V