CHAPTER

8

CONTINUOUS RANDOM VARIABLES AND THE NORMAL DISTRIBUTION

Syllabus coverage

TOPIC 4.2: CONTINUOUS RANDOM VARIABLES AND THE NORMAL DISTRIBUTION

General continuous random variables

4.2.1 use relative frequencies and histograms obtained from data to estimate probabilities associated with a continuous random variable

4.2.2 examine the concepts of a probability density function, cumulative distribution function, and probabilities associated with a continuous random variable given by integrals; examine simple types of continuous random variables and use them in appropriate contexts

4.2.3 identify the expected value, variance and standard deviation of a continuous random variable and evaluate them using technology

4.2.4 examine the effects of linear changes of scale and origin on the mean and the standard deviation

Normal distributions

4.2.5 identify contexts, such as naturally occurring variation, that are suitable for modelling by normal random variables

4.2.6 identify features of the graph of the probability density function of the normal distribution with mean μ and standard deviation σ and the use of the standard normal distribution

4.2.7 calculate probabilities and quantiles associated with a given normal distribution using technology, and use these to solve practical problems

8.1 General continuous random variables

8.1

Recall that in Chapter 5, we defined a random variable, X, as a set of numerical quantities with elements defined as the outcomes of a random chance experiment. However, there are two types of random variables: discrete and continuous. Chapter 5 dealt with discrete random variables, whose possible values are any countable set of numbers, such as the number of pets in a household. In this chapter, we will explore continuous random variables.

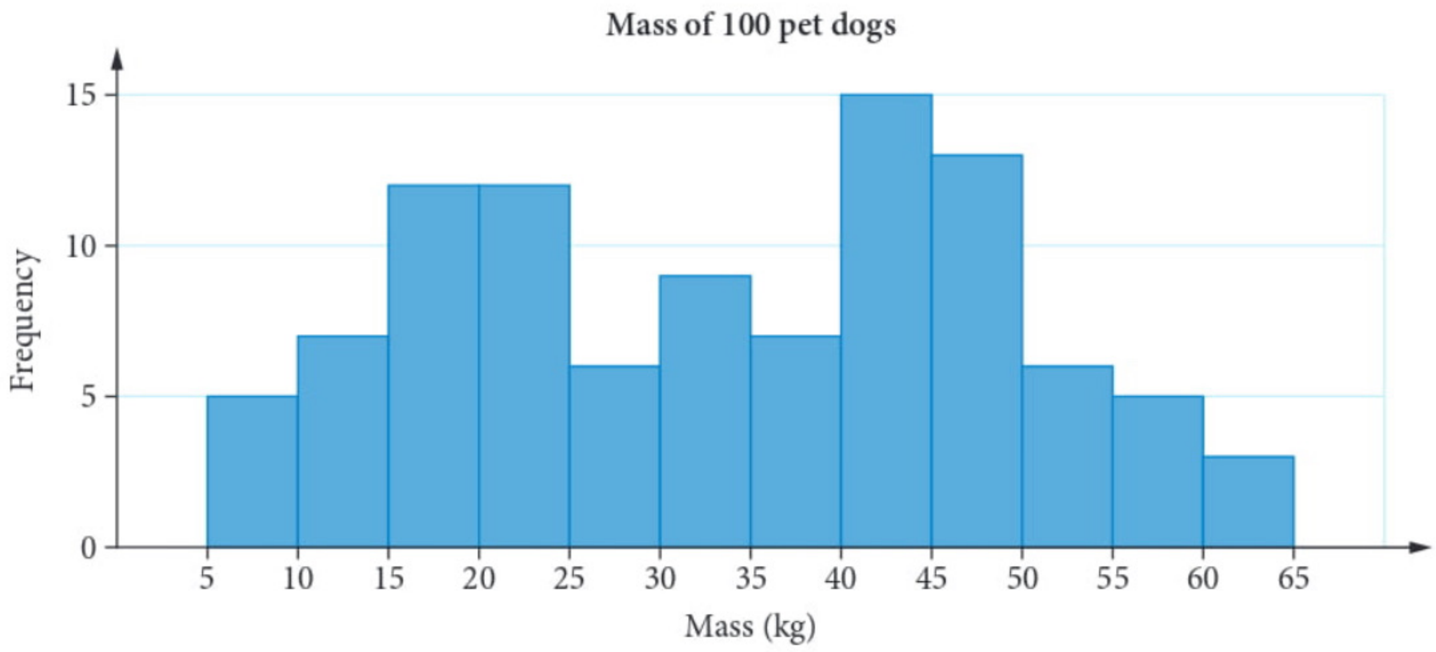
A continuous random variable can take the value of any real number over a given continuous interval. Examples of continuous random variables include the height of students in a class, the minimum daily temperatures or the mass of pet dogs.

Relative frequencies, estimates and histograms

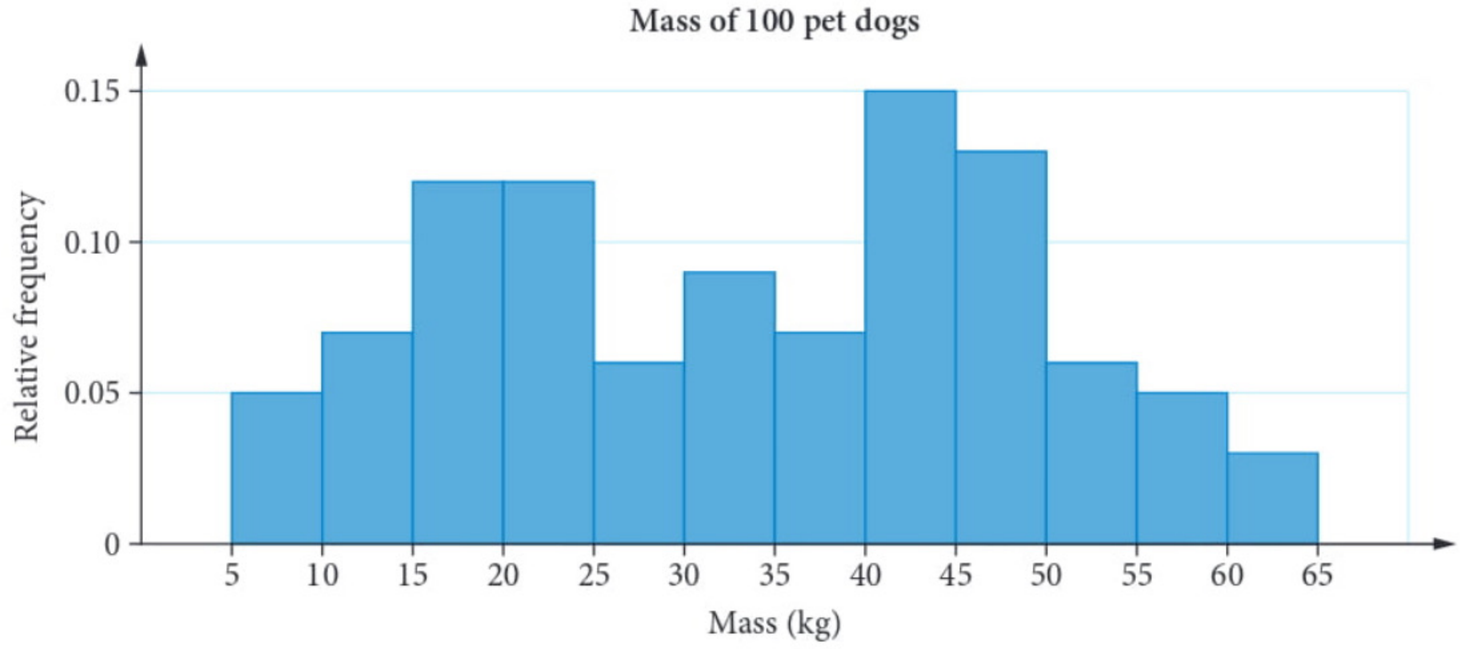
Video playlist General continuous random variables

Worksheet Probability density functions

As with discrete random variables, the idea of a continuous random variable is not new to you. You may remember constructing **frequency histograms** to display the frequencies of the outcomes of a continuous numerical data set in which the variable being measured is on the horizontal axis and the frequency is on the vertical axis. For example, suppose a local vet recorded the mass of the next 100 pet dogs that visited the clinic and displayed the results in a histogram, such as the one below.



In this example, the continuous random variable is the mass of a pet dog visiting the clinic, recorded in kilograms. Let this variable be M. From this experimental data, we can then calculate a **relative frequency** of an outcome, which is its frequency as a proportion of the total number of observations. This relative frequency is also called an **experimental probability**. Given that the probability of an outcome is coming from the collection of data, it cannot be considered a theoretical probability as we would expect there to be differences in the data collected with every set of experiments conducted. As a result, based on this single set of data, we can use the relative frequencies of each outcome as estimates of the theoretical probabilities and then carry out calculations involving the same chance experiment. For example, we could estimate that 15 out of every 100 pet dogs that visit this particular vet clinic have a mass between 40 and 45 kilograms. This can be written using the following probability notation $P(40 \le M < 45) = \frac{15}{100} = 0.15$ and all relative frequencies could be displayed in a **relative frequency histogram**.



Now without the specific dataset, we do not know whether this interval with a relative frequency of $\frac{15}{100}$ is actually $40 \le M < 45$, $40 \le M \le 45$, $40 < M \le 45$, or even $40 < M < 45$. This brings up an important question for continuous random variables:

What is the probability that a pet dog has a mass of EXACTLY 40 kg?

Or similarly,

What is the probability that a pet dog has a mass of EXACTLY 45 kg?

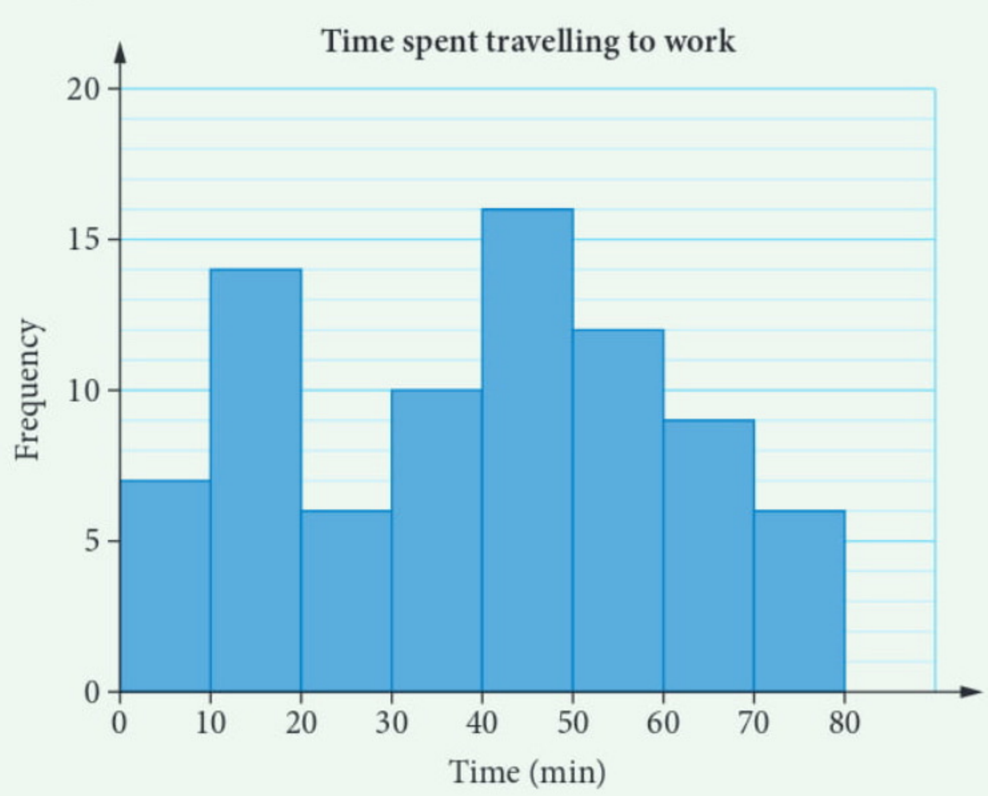
Because continuous random variables can take on any real numerical value, there is an underlying assumption that the probability that the mass of a dog is an exact discrete value such as 45 kg is negligible; that is, $P(M = 45) = 0$ because the degree of accuracy of mass can never be measured exactly. It could be 45.0001 kg or 44.999 987 4 kg.

Probability at a point for a continuous random variable

For a continuous random variable, X, $P(X = k) = 0$, where k is a discrete outcome.

WORKED EXAMPLE 1 Probabilities from a histogram

The histogram below shows the results of a survey of 80 workers who were asked how long it took them to arrive at work that morning.



Let T be the continuous random variable representing the time taken to arrive at work, measured in minutes.

a Complete the frequency table below.

Time	Frequency
$0 \le t < 10$	7
$10 \le t < 20$	14
$20 \le t < 30$	6
$30 \le t < 40$	
$40 \le t < 50$	16
$50 \le t < 60$	12
$60 \le t < 70$	9
$70 \le t < 80$	
Total	80

b Use the frequencies from this data set to estimate the following probabilities:

i $P(10 \le T < 20)$ **ii** $P(T \ge 10 \mid T \le 20)$ **iii** $P(T < 50)$

iv $P(T \ge 30)$ **v** $P(25 \le T \le 45)$

Steps	Working
a **1** Read the frequencies from the histogram to complete the table.	$n(30 \le T < 40) = 10$ $n(70 \le T < 80) = 6$
2 Check that the frequencies total 80.	$7 + 14 + 6 + 10 + 16 + 12 + 9 + 6 = 80$
b **i** Express the frequency for $10 \le t < 20$ as a relative frequency and use it as an estimate of the probability.	$P(10 \le T < 20) = \frac{14}{80}$
ii Use the conditional probability formula $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ and recognise that $P(T \le 20) = P(T < 20)$ for the continuous random variable.	$P(T \ge 10 \mid T \le 20) = \frac{P(10 \le T < 20)}{P(T < 20)}$ $= \frac{14}{21}$
iii Add up all frequencies for $t < 50$ and express the **cumulative relative frequency** as an estimate of the probability.	$P(T < 50) = \frac{7 + 14 + 6 + 10 + 16}{80}$ $= \frac{53}{80}$
iv Use the complement rule to $P(T \ge 30) = 1 - P(T < 30)$ to obtain an estimate of the probability.	$P(T \ge 30) = 1 - \frac{27}{80}$ $= \frac{53}{80}$

Exam hack

Unless you are specifically asked to simplify a fraction in the exam, there's no need to.

v 1 Assume a uniform distribution of the times within each interval to estimate the proportions from $20 \le T < 30$ and $40 \le T < 50$.

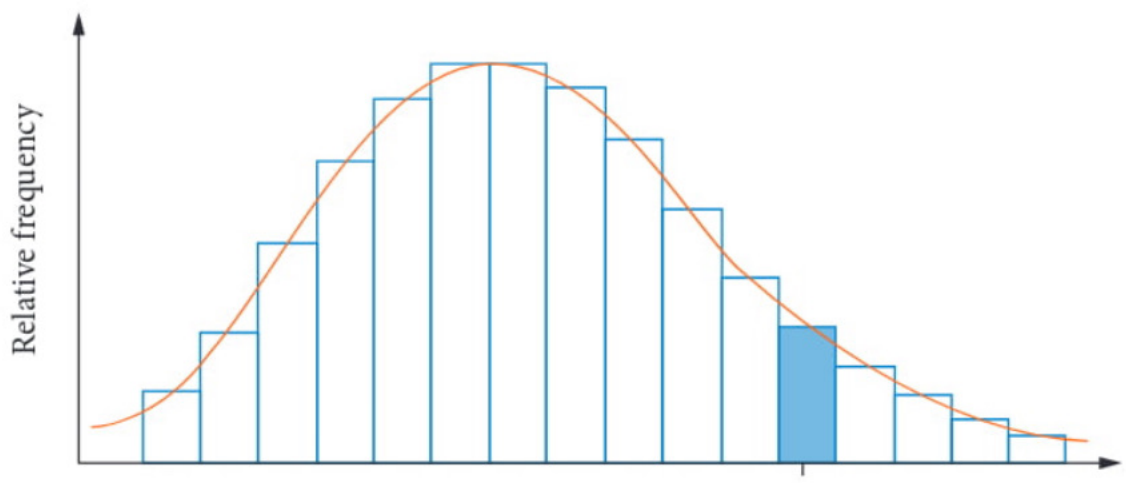
$$P(25 \le T < 30) \approx \frac{5}{10} \times \frac{6}{80}$$
$$\approx \frac{3}{80}$$
$$P(40 \le T < 45) \approx \frac{5}{10} \times \frac{16}{80}$$
$$\approx \frac{8}{80}$$

2 Express the cumulative relative frequency as an estimate of the probability.

$$P(25 \le T \le 45) \approx \frac{3 + 10 + 8}{80}$$
$$\approx \frac{21}{80}$$

Probability density functions and integrals

Suppose the vet clinic collected the mass of pet dogs over an entire year. With significantly more data, the width of the columns would become more precise and there would be significantly more columns such that the tops of the columns would start to form a smooth curve representing the range of relative frequencies over the entire interval.

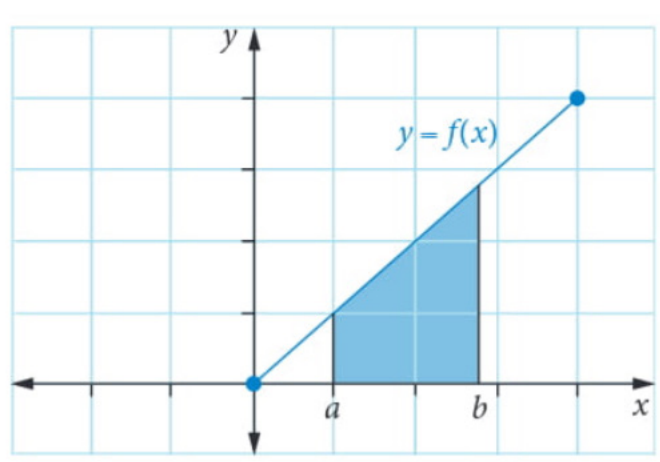


This smooth curve that is formed, which can also sometimes be defined by a rule $f(x)$, is called the **probability density function (pdf)** of a continuous random variable, X. Given that the area under the curve represents the columns of a relative frequency histogram with infinitely small width, we can say that the probability of an interval $a \le X \le b$ is given by the definite integral of the probability density function $f(x)$.

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Based on the earlier result that $P(X = k) = 0$, then

$$P(a \le X \le b) = P(a < X < b) = P(a < X \le b) = P(a \le X < b) = \int_a^b f(x)\,dx.$$



Much like the two conditions of the probability distribution of a discrete random variable, there are two required properties of a probability density function of a continuous random variable for it to be considered valid:

1 $f(x) \ge 0$ for all values of x.

2 The total area under the curve from the lowest value of x to the highest value of x is 1. This is often represented as the definite integral, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, where $\pm\infty$ represent the lower and upper bounds of x.

WORKED EXAMPLE 2 **Finding an unknown in $f(x)$ for a valid continuous random variable**

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} kx & 1 \le x \le 7 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k.

Steps	Working
1 Establish a definite integral representing the total area under the curve.	$\int_1^7 kx\,dx = 1$
2 Evaluate the definite integral and solve for k.	$\left[\frac{kx^2}{2}\right]_1^7 = 1$ $\frac{49k}{2} - \frac{k}{2} = 1$ $24k = 1$ $k = \frac{1}{24}$

In some simple cases, like the one above, an integral calculation may not be required, as the area under the curve could be calculated using simpler area formulas, such as:

area of a rectangle = lw

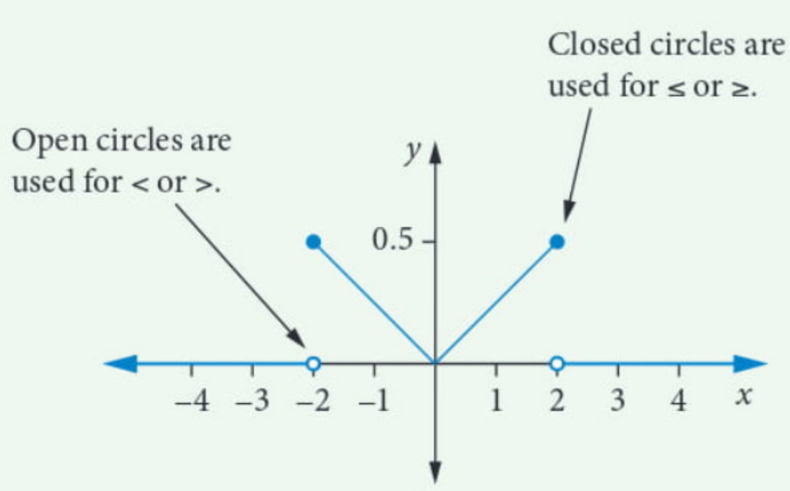
area of a triangle = $\frac{1}{2}bh$

area of a trapezium = $\frac{1}{2}(a+b)h$

area of a sector = $\frac{\theta}{360^\circ}\pi r^2$

WORKED EXAMPLE 3 **Probabilities from the graph of a probability density function**

The graph of the probability density function $f(x)$ for a continuous random variable X is shown below.



Determine the following probabilities:

a $P(X = 1)$ **b** $P(0 \le X \le 1)$ **c** $P(-1 < X < 1)$ **d** $P(X > -1 \mid X < 1)$

Steps	Working
a Recognise that $P(X = k) = 0$.	$P(X = 1) = 0$
b Use the area of a triangle formula to calculate the probability, interpreting the height of the triangle using the gradient of the line.	$P(0 \le X \le 1) = \frac{1}{2}(1)\left(\frac{1}{4}\right)$ $= \frac{1}{8}$

c	**1** Use the symmetry of the diagram to recognise that $P(0 \le X \le 1) = P(-1 \le X \le 0)$.	$P(-1 < X < 1) = 2\left(\frac{1}{8}\right)$
	2 Recognise that $P(X = k) = 0$ and so $P(-1 \le X \le 1) = P(-1 < X < 1)$.	$= \frac{1}{4}$
d	**1** Use the conditional probability formula $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.	$P(X > -1 \mid X < 1)$ $= \frac{P(-1 < X < 1)}{P(X < 1)}$
	2 Recognise that $P(X \le 1) = P(X < 0) + P(0 \le X < 1)$.	$= \frac{\frac{1}{4}}{\frac{1}{2} + \frac{1}{8}}$ $= \frac{\left(\frac{2}{8}\right)}{\left(\frac{5}{8}\right)} = \frac{2}{5}$

Exam hack

In cases where the graph of the probability density function is not given to you, if you recognise the shape of the graph from the rule, always draw a quick sketch of it to help represent the bounds of the definite integral.

WORKED EXAMPLE 4 Probabilities from the rule of a probability density function

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{3}{32}x(4 - x) & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Determine the following probabilities:

a $P(X < 4)$ **b** $P(0 \le X \le 2)$ **c** $P(X \ge 3)$ **d** $P(X > 0 \mid X < 2)$

Steps	**Working**
a Recognise that all values of x are less than 4 and that $P(X = 4) = 0$.	$P(X < 4) = P(0 \le X \le 4) = 1$
b **1** Establish a definite integral for the area under the curve.	$\int_0^2 \frac{3}{32}x(4 - x)\,dx$ $= \frac{3}{32}\int_0^2 4x - x^2\,dx$
2 Evaluate the definite integral.	$= \frac{3}{32}\left[2x^2 - \frac{x^3}{3}\right]_0^2$ $= \frac{3}{32}\left(8 - \frac{8}{3}\right)$ $= \frac{3}{32}\left(\frac{16}{3}\right)$ $= \frac{1}{2}$

Exam hack

If you can recognise critical features of a probability density function such as this function having roots at $x = 0$ and $x = 4$, with a line of symmetry at $x = 2$, then pay attention to the number of marks. For 2 marks or less, it could be inferred that $P(0 \le X \le 2) = \frac{1}{2}$.

c Recognise the probability statement has an upper bound at 4 and establish the definite integral.

$$P(3 \leq X < 4) = \int_3^4 \frac{3}{32}x(4-x)\,dx$$
$$= \frac{3}{32}\left[2x^2 - \frac{x^3}{3}\right]_3^4$$
$$= \frac{3}{32}\left(32 - \frac{64}{3} - 18 + 9\right)$$
$$= 1 - \frac{27}{32} = \frac{5}{32}$$

d **1** Use the conditional probability formula $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

$$P(X > 0 \mid X < 2) = \frac{P(0 < X < 2)}{P(X < 2)}$$

2 Recognise that $P(0 < X < 2) = P(0 \leq X < 2)$ (from part **b**) and $P(X < 2) = P(X \leq 2)$.

$$= \frac{\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right)} = 1$$

Exam hack

Questions such as this could be worth 1 mark. If a conditional probability question is only allocated 1 mark, there will usually be something to notice about the intervals given. For example, in this case it is certain that $x > 0$ if $x < 2$ as the function starts from $x = 0$.

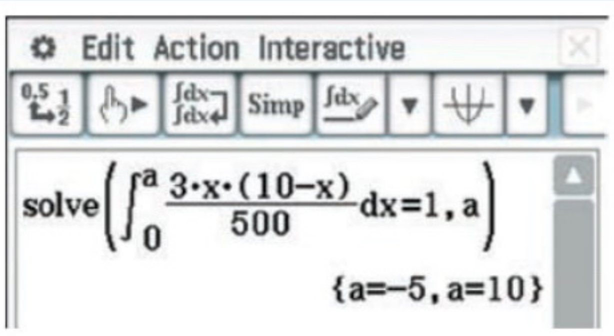
USING CAS 1 Finding an unknown in the domain of $f(x)$ for a valid continuous random variable

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{3x(10-x)}{500} & \text{if } 0 \leq x \leq a \\ 0 & \text{otherwise} \end{cases}$$

Find the value of a.

ClassPad

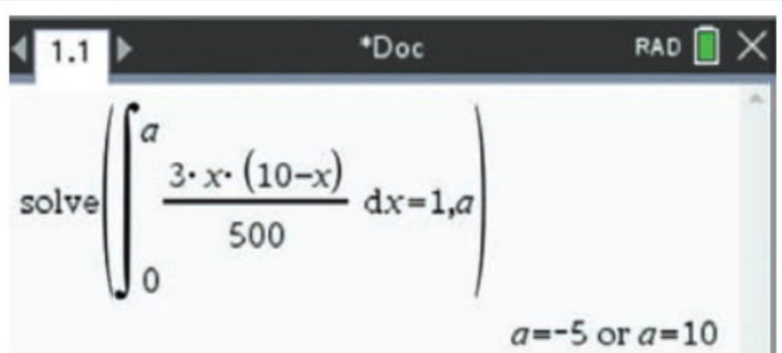


1 Enter the definite integral with lower and upper limits of **0** and **a** and set equal to **1**.

2 Highlight the integral equation and solve for a, as shown above.

3 Select the solution that suits the domain, i.e. in this case greater than 0.

TI-Nspire



1 Set the definite integral equal to 1 and solve for **a**, as shown above.

2 Select the positive solution.

$a \neq -5$ as this is outside the domain.

$a = 10$

Exam hack

In the Calculator-assumed section, if you can see that a probability density function is being used multiple times throughout a question, it may be beneficial for time efficiency to use CAS to define the probability density function. It is especially helpful when the probability calculation goes across two separate functions in the piece-wise definition.

USING CAS 2 Defining probability density functions

Consider the following probability density function $f(x)$ for the continuous random variable X.

$$f(x) = \begin{cases} \frac{x}{4} & 0 \le x < 2 \\ 1 - \frac{x}{4} & 2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

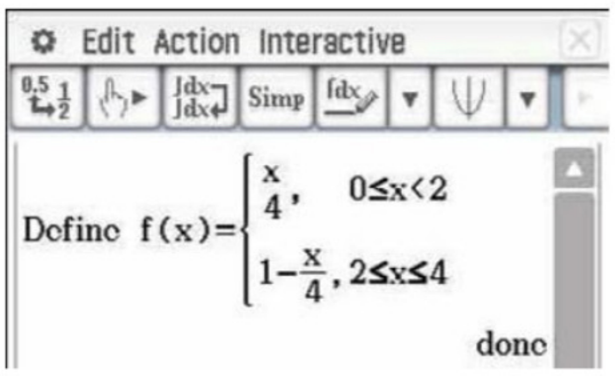
Calculate the following probabilities:

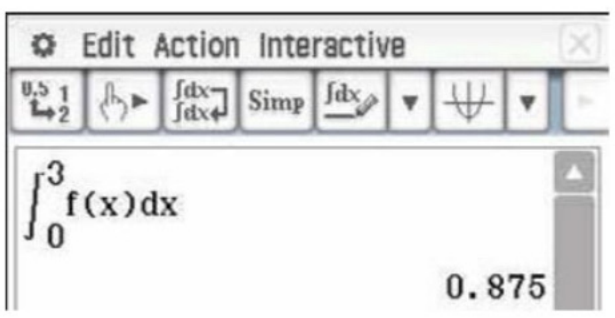
a $P(X < 3)$ **b** $P(X \ge 1 \mid X < 2)$ **c** $P(X > 3.5 \mid X > 1.5)$

ClassPad | **TI-Nspire**

a

ClassPad:

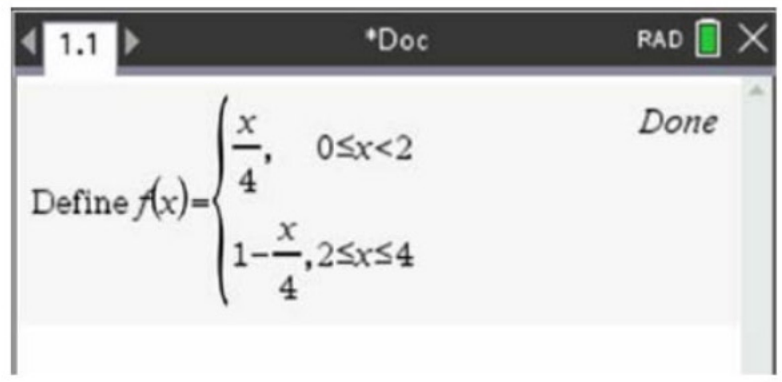




1 Using the **Math3** keyboard, **Define f(x)** piece-wise-defined function template.

2 Find the definite integral of **f(x)** from 0 to 3.

TI-Nspire:

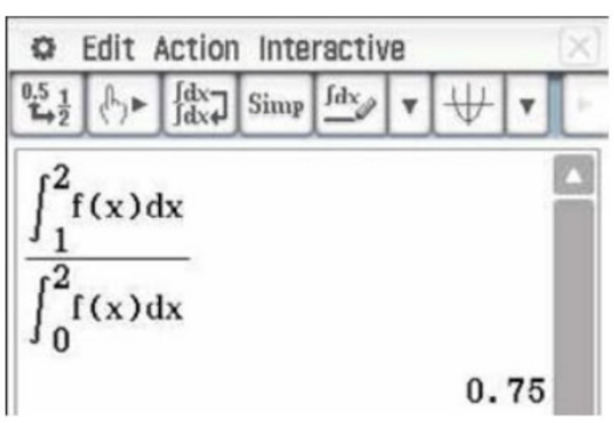


1 Define the piece-wise function **f(x)**.

2 Find the definite integral of **f(x)** from 0 to 3.

b

ClassPad:



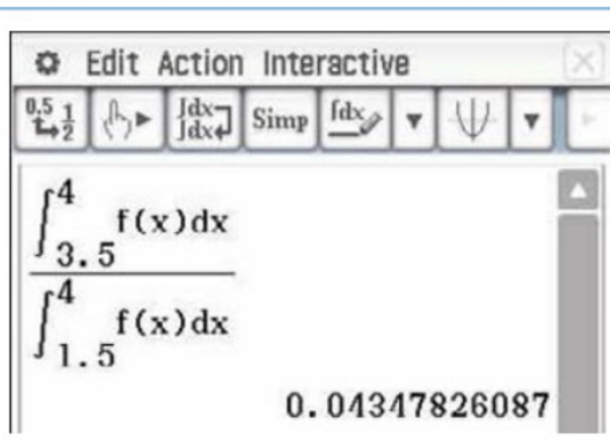
Use the conditional probability formula to find the definite integral of **f(x)** from 1 to 2, divided by the definite integral of **f(x)** from 0 to 2.

TI-Nspire:

Use the conditional probability formula to find the definite integral of **f(x)** from 1 to 2, divided by the definite integral of **f(x)** from 0 to 2.

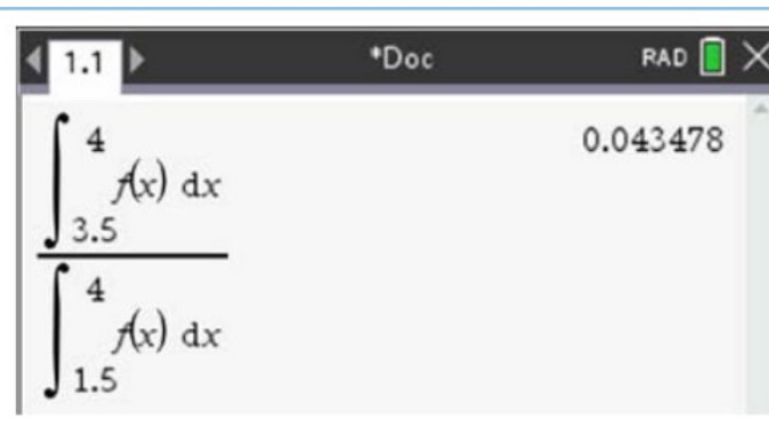
c

ClassPad:



Use the conditional probability formula* to find the definite integral of **f(x)** from 3.5 to 4, divided by the definite integral of **f(x)** from 1.5 to 4.

TI-Nspire:



Use the conditional probability formula* to find the definite integral of **f(x)** from 3.5 to 4, divided by the definite integral of **f(x)** from 1.5 to 4.

a 0.875 **b** 0.75 **c** 0.0435

*Note: In this case $P(A \cap B) = P(A)$ because the interval $X > 3.5$ (i.e. set A) is completely contained within the interval $X > 1.5$ (i.e. set B).

Cumulative probability distributions

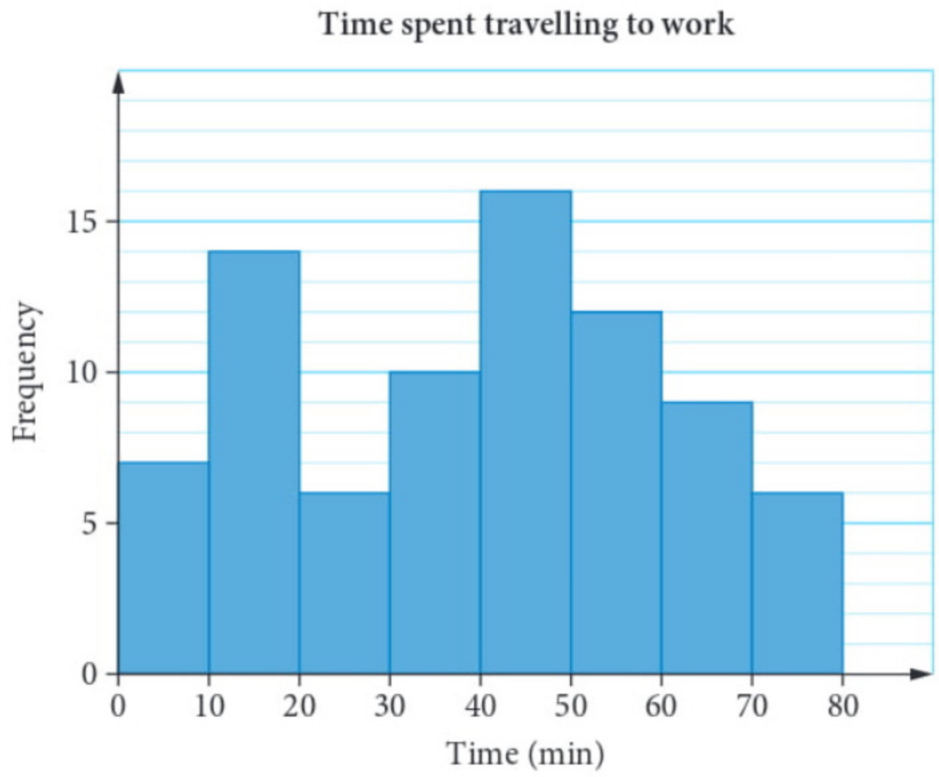
8.1

In many of the calculations so far, we have had to use the idea of a cumulative probability; that is, the addition of probabilities for a continuous random variable X from consecutive intervals up to a certain value of x, $P(X \leq x)$. We have also seen that $P(X > x)$ can also be expressed as a cumulative probability using the complement rule:

$$P(X > x) = 1 - P(X \leq x)$$

Now suppose we wanted to expressed the cumulative probabilities generally, either as a table of values, graphically or as a function.

Recall the histogram and corresponding frequency table that showed the results of a survey of 80 workers who were asked how long it took them to arrive at work that morning, where T was the continuous random variable representing the time taken to arrive at work, measured in minutes.



Time	Frequency
$0 \leq t < 10$	7
$10 \leq t < 20$	14
$20 \leq t < 30$	6
$30 \leq t < 40$	10
$40 \leq t < 50$	16
$50 \leq t < 60$	12
$60 \leq t < 70$	9
$70 \leq t < 80$	6
Total	80

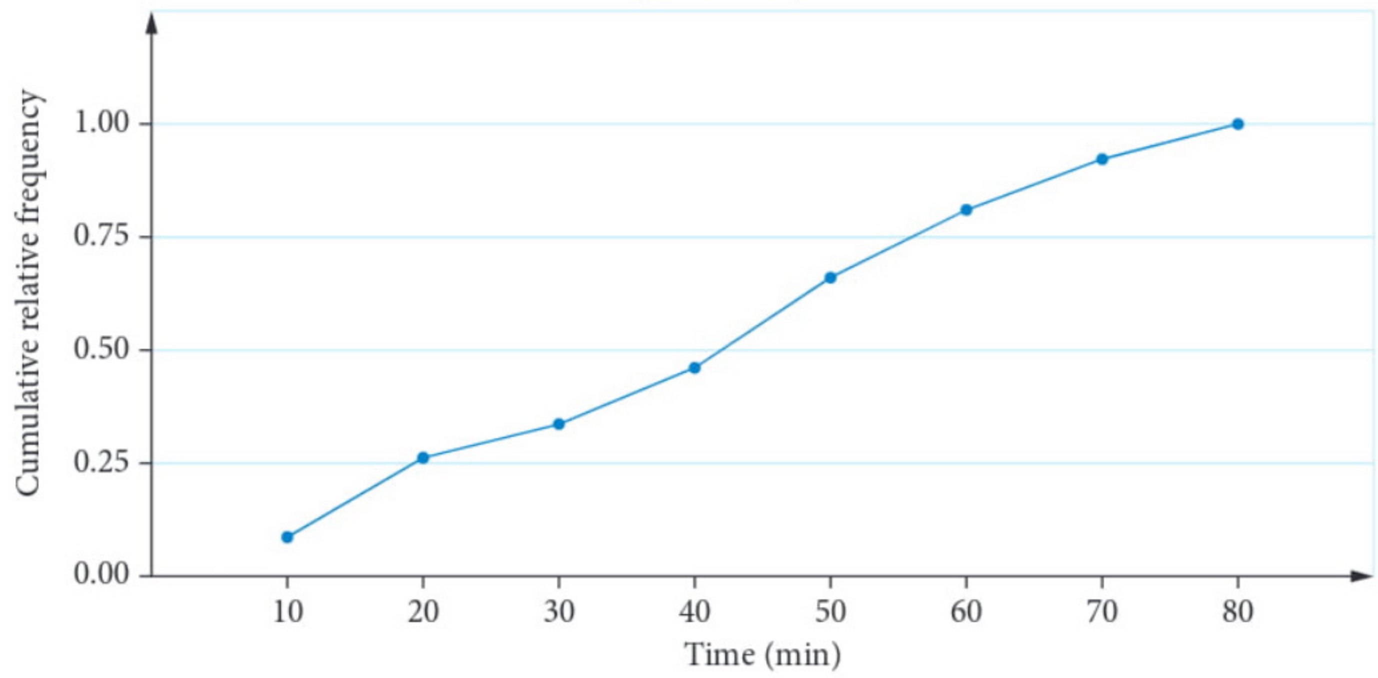
Instead of displaying this data in a frequency table, we can display it in a cumulative frequency table, which then makes calculations of the form $P(T \geq t)$ easier.

Time	Cumulative frequency	Cumulative relative frequency
$t < 10$	7	$\frac{7}{80}$
$t < 20$	21	$\frac{21}{80}$
$t < 30$	27	$\frac{27}{80}$
$t < 40$	37	$\frac{37}{80}$
$t < 50$	53	$\frac{53}{80}$
$t < 60$	65	$\frac{65}{80}$
$t < 70$	74	$\frac{74}{80}$
$t < 80$	80	$\frac{80}{80}$
Total	80	1

Note that the final row of the cumulative frequency column should give the same value as the total, and the final row of the cumulative relative frequency column should give 1.

Suppose now that this cumulative data was used to create a smooth curve through the tops of the cumulative frequency histogram columns. We would obtain a function such as the one below, with the horizontal axis representing T and the vertical axis representing the cumulative relative frequencies whereby $P(T < 80) = 1$. This is called the **cumulative probability distribution** of T.

Cumulative probability distribution of *T*



In cases where a rule for the function of the cumulative probability distribution can be obtained, then it is called the **cumulative distribution function** **(cdf)** and typically denoted using $F(x)$.

$$P(X \leq x) = F(x)$$

For a continuous random variable X with a probability density function f defined over $a \leq x \leq b$, then the cumulative distribution function F is summing all areas under the curve of f from $x = a$ to some value of x. Let f be defined in terms of the dummy variable t, and so using the fundamental theorem of calculus, we can say that $F(x) = \int_a^x f(t)\,dt$ and so $\frac{d}{dx}(F(x)) = \frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x)$.

Cumulative distribution function

The probability density function $f(x)$ of a continuous random variable X is the derivative of the cumulative distribution function $F(x)$ and, hence, the cumulative distribution function $F(x)$ is the integral of $f(x)$ such that

$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x f(t)\,dt & a \leq x \leq b \\ 1 & x > b \end{cases}$$

WORKED EXAMPLE 5 Finding a cumulative distribution function given a probability density function

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{48} & 2 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

a Determine the cumulative distribution function of X.

b Hence, calculate $P(X \leq 5)$.

Steps	Working
a **1** Replace the x in $f(x)$ with a dummy variable t and establish the definite integral $F(x) = \int_a^x f(t)\,dt$.	$f(t) = \dfrac{t}{48}$
2 Evaluate the definite integral in terms of x and establish the piece-wise definition of $F(x)$.	$F(x) = \int_2^x \dfrac{t}{48}\,dt$ $= \left[\dfrac{t^2}{96}\right]_2^x$ $= \dfrac{x^2 - 4}{96}$ $F(x) = \begin{cases} 0 & x < 2 \\ \dfrac{x^2-4}{96} & 2 \le x \le 10 \\ 1 & x > 10 \end{cases}$
b Evaluate $F(5)$.	$F(5) = \dfrac{5^2 - 4}{96} = \dfrac{21}{96}$

Exam hack

When a probability density function has more than two components in its piece-wise definition, write out the addition of the definite integrals that would give you a total area under the curve of 1.

WORKED EXAMPLE 6 Finding a probability density function given a cumulative distribution function

The cumulative distribution function for a continuous random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^2}{8} & 0 \le x < 2 \\ x - \dfrac{x^2}{8} - 1 & 2 \le x \le 4 \\ 1 & x > 4 \end{cases}$$

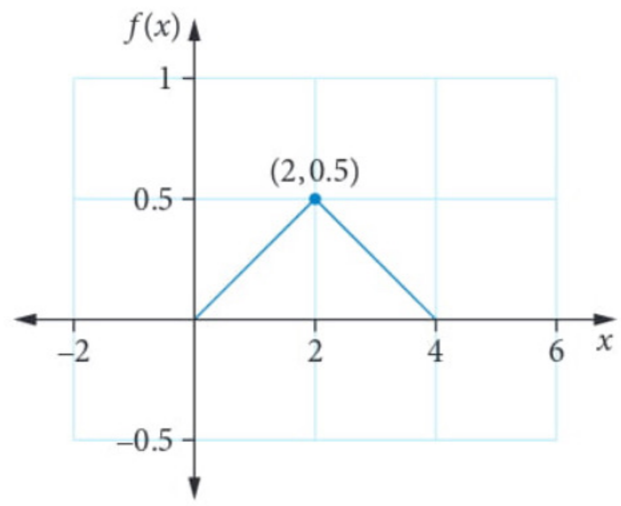
a Find $P(X > 3)$.

b Determine the probability density function of X, $f(x)$.

c Sketch the graph of $f(x)$.

d Hence, show how to obtain the rule $F(x) = x - \dfrac{x^2}{8} - 1$ for $2 \le x \le 4$ from $f(x)$.

Steps	Working
a **1** Express $P(X > 3)$ in terms of a cumulative probability.	$P(X > 3) = 1 - P(X \le 3)$
2 Use $F(x)$ to evaluate the probability.	$P(X > 3) = 1 - F(3)$ $= 1 - \left(3 - \dfrac{9}{8} - 1\right)$ $= 1 - \dfrac{7}{8}$ $= \dfrac{1}{8}$

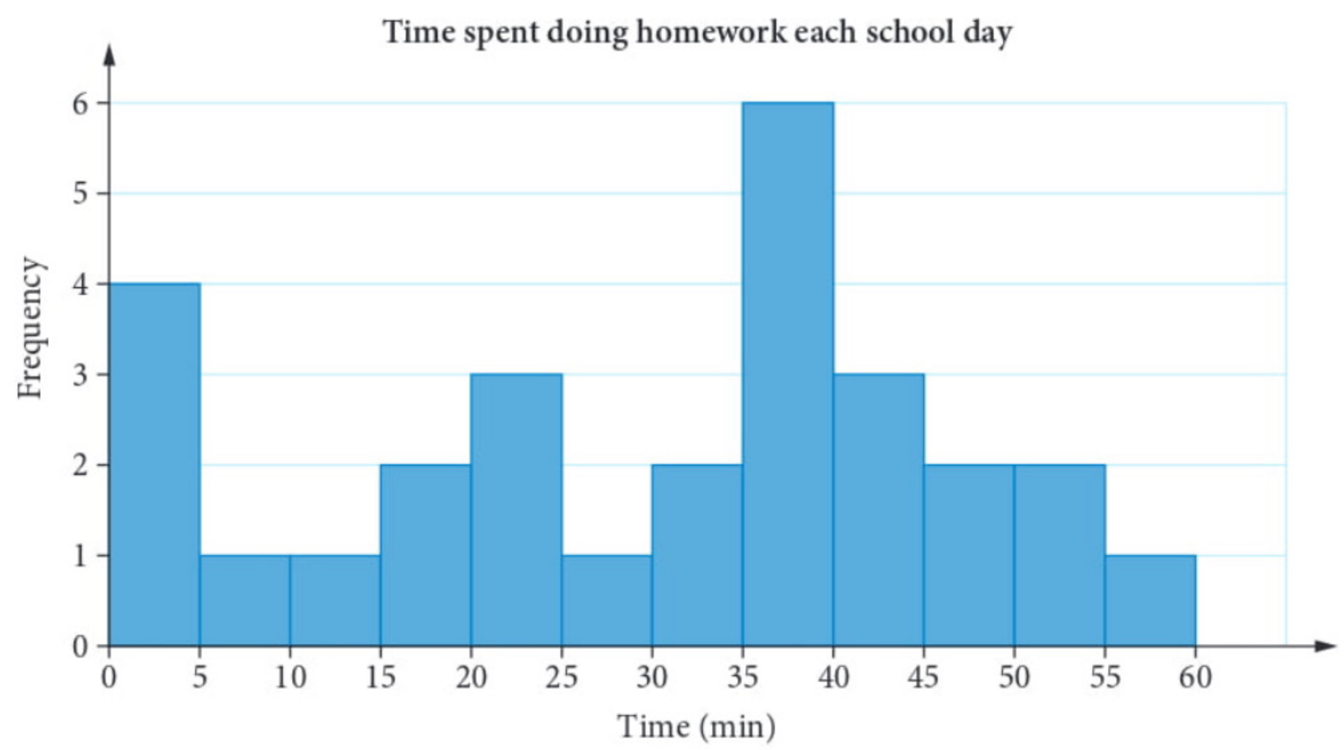
b **1** Obtain $f(x)$ by differentiating each component of $F(x)$.	For $0 \le x < 2$, $f(x) = \frac{d}{dx}\left(\frac{x^2}{8}\right) = \frac{x}{4}$ For $2 \le x \le 4$, $f(x) = \frac{d}{dx}\left(x - \frac{x^2}{8} - 1\right) = 1 - \frac{x}{4}$
2 Express $f(x)$ in the piece-wise definition.	$f(x) = \begin{cases} \frac{x}{4} & 0 \le x < 2 \\ 1 - \frac{x}{4} & 2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$
c **1** Find the end points of each component of the piece-wise function.	$f(0) = 0$ $f(2) = \frac{1}{2}$ $f(4) = 0$
2 Sketch the graph of $y = f(x)$ ensuring all critical features are shown.	
d **1** Establish an expression using definite integrals for the area under the curve of $f(x)$ from 0 to any value of x between $2 \le x \le 4$.	For $2 \le x \le 4$, $F(x) = \int_0^2 \frac{x}{4}\,dx + \int_2^x 1 - \frac{t}{4}\,dt$ $= \frac{1}{2}(2)\left(\frac{1}{2}\right) + \left[t - \frac{t^2}{8}\right]_2^x$
2 Use geometry, where efficient, to assist with any calculations.	$= \frac{1}{2} + \left(x - \frac{x^2}{8} - 2 + \frac{1}{2}\right)$ $= x - \frac{x^2}{8} - 1$

EXERCISE 8.1 General continuous random variables

ANSWERS p. 403

Mastery

1 WORKED EXAMPLE 1 The histogram below shows the amount of time, in minutes, that Jared spends completing homework each day, over the period of 28 school days. He does some homework on every school day.



Let T be the continuous random variable representing the time spent completing homework on a school day, measured in minutes.

a Copy and complete the frequency table below.

Time	Frequency
$0 \le t < 5$	4
$5 \le t < 10$	1
$10 \le t < 15$	1
$15 \le t < 20$	2
$20 \le t < 25$	
$25 \le t < 30$	1
$30 \le t < 35$	2
$35 \le t < 40$	
$40 \le t < 45$	3
$45 \le t < 50$	2
$50 \le t < 55$	2
$55 \le t < 60$	1
Total	28

b Use the frequencies from this data set to estimate the following probabilities:

i $P(15 \le T \le 20)$ **ii** $P(T \ge 15 \mid T \le 20)$ **iii** $P(T < 40)$

iv $P(T \ge 20)$ **v** $P(42 \le T \le 48)$

2 WORKED EXAMPLE 2 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} 5kx - kx^2 & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k.

3 WORKED EXAMPLE 3 The graph of the probability density function $f(x)$ for a continuous random variable X is shown.

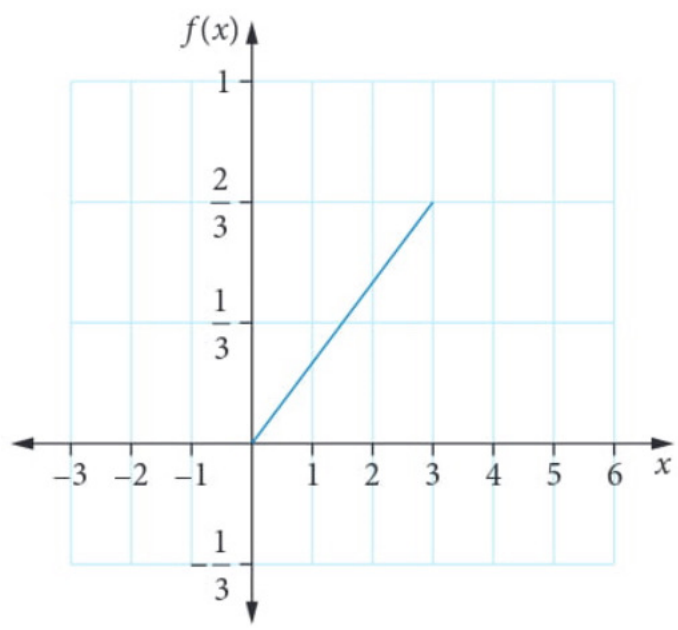
Determine the following probabilities:

a $P(X = 2)$

b $P(0 \le X \le 2)$

c $P(1 < X < 2)$

d $P(X > 1 \mid X < 2)$



4 WORKED EXAMPLE 4

a The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} 4x^3 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine $P\left(X > \frac{1}{2}\right)$.

b The probability density function for a continuous random variable Y is given by

$$f(y) = \begin{cases} 0.01e^{-0.01y} & y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

Determine $P(50 \le Y \le 80)$, leaving your answer exact.

c The probability density function for a continuous random variable Z is given by

$$f(z) = \begin{cases} \pi \sin(2\pi z) & 0 \le z \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Determine $P\left(Z > \frac{1}{4} \mid Z < \frac{1}{3}\right)$, leaving your answer exact.

5 Using CAS 1 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{4x^3}{625} & 0 \le x \le a \\ 0 & \text{otherwise} \end{cases}$$

Find the value of a.

6 Using CAS 2 Consider the following probability density function $f(x)$ for the continuous random variable X.

$$f(x) = \begin{cases} \frac{5-x}{25} & 0 \le x < 5 \\ \frac{x-5}{25} & 5 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the following probabilities:

a $P(X < 8)$ **b** $P(X \ge 2 \mid X < 5)$ **c** $P(X > 8 \mid X > 5)$

7 WORKED EXAMPLE 5 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{2x}{9} & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

a Determine the cumulative distribution function of X.

b Hence, calculate $P\left(X \le \frac{5}{2}\right)$.

8 WORKED EXAMPLE 6 The cumulative distribution function for a continuous random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{4} - \frac{x^2}{32} & 0 \le x < 4 \\ \frac{x^2}{32} - \frac{x}{4} + 1 & 4 \le x \le 8 \\ 1 & x > 8 \end{cases}$$

a Find $P(X \ge 6)$.

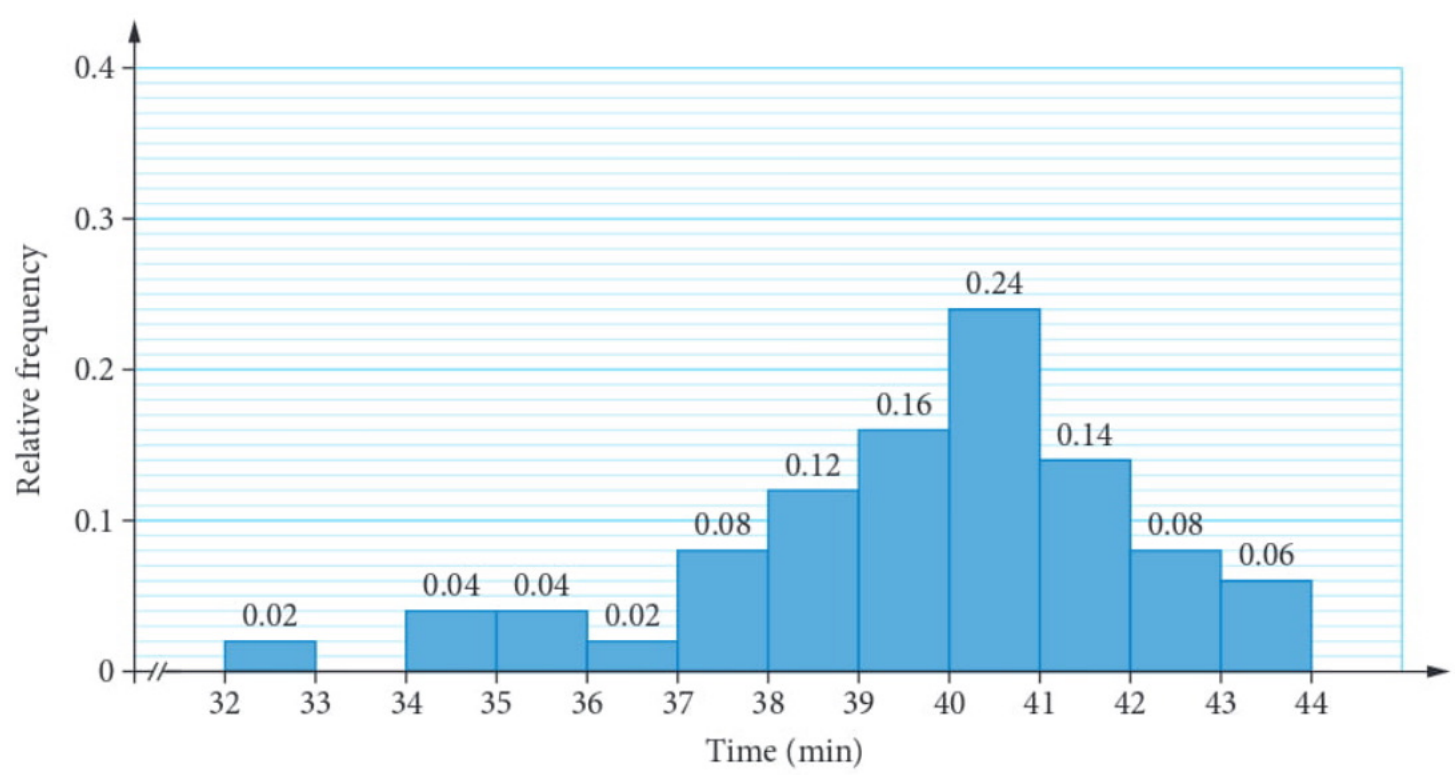
b Determine the probability density function of X, $f(x)$.

c Sketch the graph of $f(x)$.

d Hence, show how to obtain the rule $F(x) = \frac{x^2}{32} - \frac{x}{4} + 1$ for $4 \le x \le 8$ from $f(x)$.

Calculator-free

9 © SCSA MM2017 Q1 (5 marks) Anastasia is a university student. She records the time it takes for her to get from home to her campus each day. The histogram of relative frequencies below shows the journey times she recorded.



Use the above data to estimate the probability of her next journey from home to her university campus

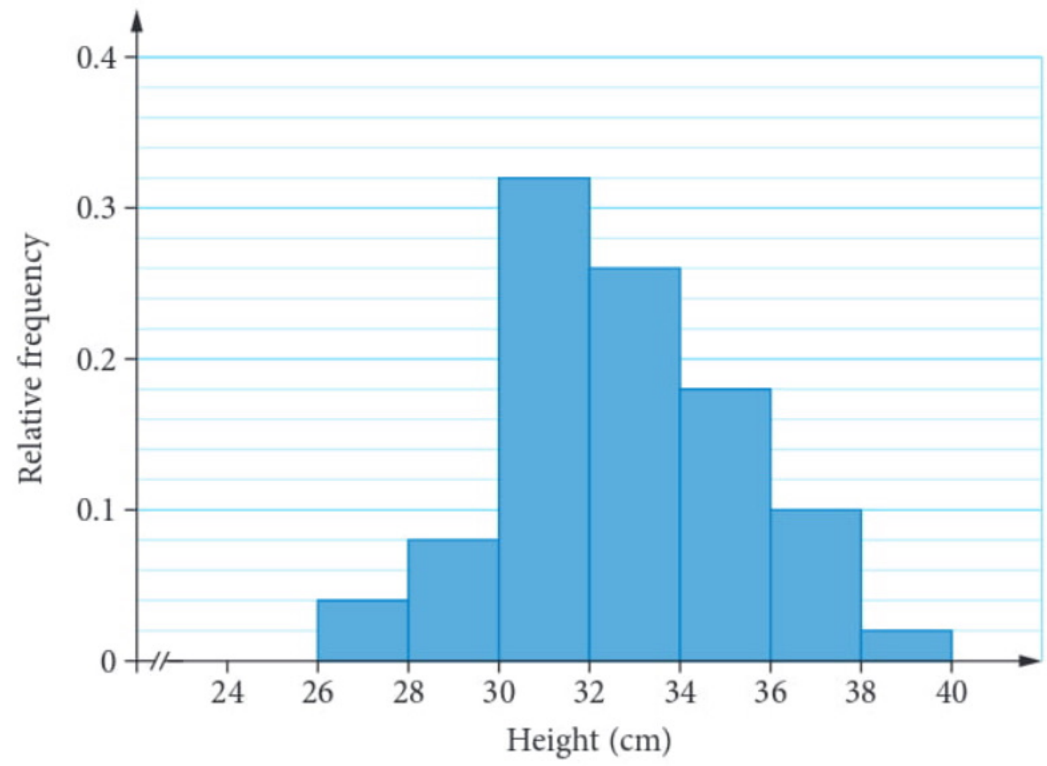
a taking her less than 36 minutes (1 mark)

b taking at least 35 minutes but no more than 39 minutes. (2 marks)

On three consecutive days, Anastasia needs to be on campus no later than 10 am.

c If she leaves her home at 9:22 am each day, use the above data to estimate the probability that she makes it on or before time on all three days. (2 marks)

10 © SCSA MM2020 Q4 (9 marks) The heights reached by a species of small plant at maturity are measured by a team of biologists. The results are shown in the histogram of relative frequencies below.



a Determine the probability that a mature plant of this species reaches no higher than 30 cm. (1 mark)

b If a mature plant reaches a height of at least 32 cm, what is the probability that its height reaches above 38 cm? (2 marks)

Another team of biologists is studying the mature heights of a species of hedge. The height, h metres, has a probability density function, $d(h)$, as given below.

$$d(h) = \begin{cases} \dfrac{h-1}{5} & 1 \leq h \leq 2 \\ kh^2 & 2 < h \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

c What percentage of hedges from this study reaches a mature height less than 2 m? (3 marks)

d Determine the value of k. (3 marks)

11 (7 marks) The probability density function $f(x)$ of a continuous random variable X is given by

$$f(x) = \begin{cases} \dfrac{x+1}{k} & 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = 12$. (2 marks)

b Find the value b of such that $P(X \leq b) = \dfrac{5}{8}$. (3 marks)

c Hence, determine $P(X > 2 \mid X < 3)$. (2 marks)

12 (4 marks) A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \dfrac{a}{x^2} & x \geq a \\ 0 & \text{otherwise} \end{cases}$$

a Use the expression $\lim\limits_{x \to \infty}\left(\int_a^k \dfrac{a}{x^2}\,dx\right)$ to justify why $f(x)$ is a valid probability density function. (2 marks)

b Determine $P(X > 2a)$. (2 marks)

9780170477536

Calculator-assumed

13 © SCSA MM2017 Q11a (2 marks) A pizza shop estimates that the time X hours to deliver a pizza from when it is ordered is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{4}{3} - \frac{2}{3}x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine the probability of a pizza being delivered within half an hour of being ordered.

14 © SCSA MM2018 Q10ab (5 marks) The following function is a probability density function on the given interval:

$$f(x) = \begin{cases} ax^2(x-2) & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Find the value of a. (3 marks)

b Find the probability that $x \ge 1.2$. (2 marks)

15 (7 marks) Each night Kim goes to the gym or the pool. When Kim goes to the gym, the time, T hours, that she spends working out is a continuous random variable with probability density function given by

$$f(t) = \begin{cases} 4t^3 - 24t^2 + 44t - 24 & \text{for } 1 \le t \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(t)$. Label any stationary points with their coordinates, correct to two decimal places. (3 marks)

b What is the probability, correct to three decimal places, that she spends less than 75 minutes working out when she goes to the gym? (2 marks)

c Kim calls her longest workout sessions 'super sessions'. If Kim spends more than k minutes in the gym she will have done a 'super session' and the probability of this occurring is 0.1. Find the value of k, to the nearest minute. (2 marks)

16 (4 marks) Sharelle is the goal shooter for her netball team. The time in hours that Sharelle spends training each day is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{1}{64}(6-x)(x-2)(x+2) & \text{for } 2 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the probability density function, and label the local maximum with its coordinates, correct to two decimal places. (2 marks)

b What is the probability, correct to four decimal places, that Sharelle spends less than 3 hours training on a particular day? (2 marks)

17 (6 marks)

a The continuous random variable, X, has a probability density function given by

$$f(x) = \begin{cases} \frac{1}{4}\cos\left(\frac{x}{2}\right) & 3\pi \le x \le 5\pi \\ 0 & \text{otherwise} \end{cases}$$

Determine the value of a to two decimal places such that $P(X < a) = \frac{\sqrt{3}+2}{4}$. (3 marks)

b A probability density function $f(t)$ for a continuous random variable T is given by

$$f(t) = \begin{cases} \cos(t) + 1 & k \le t \le (k+1) \\ 0 & \text{otherwise} \end{cases}$$

where $0 < k < 2$. Show that the exact value of k is $\frac{\pi - 1}{2}$. (3 marks)

18 (7 marks) In a chocolate factory, the time, Y seconds, taken to produce a chocolate has the following probability density function.

$$f(y) = \begin{cases} 0 & y < 0 \\ \frac{y}{16} & 0 \le y \le 4 \\ 0.25e^{-0.5(y-4)} & y > 4 \end{cases}$$

a Explain, with the appropriate working, why $f(y)$ is a valid probability density function for a continuous random variable Y. (4 marks)

b Find, correct to four decimal places, $P(3 \le Y \le 5)$. (3 marks)

8.2 Measures of centre and spread

Video playlist Measures of centre and spread

Expected value (mean) as a measure of centre

Recall from Chapter 5 that the expected value of a random variable X is the summation of all outcomes x multiplied by their corresponding probabilities $p(x)$. In discrete cases, we used the summation notation below as the values of $p(x)$ were defined for discrete values of x.

$$E(X) = \sum_{i=1}^{n} x_i p(x_i)$$

However, now that X is continuous, $p(x)$ is also a continuous curve and so the summation takes the form of a definite integral.

The expected value (mean) of a continuous random variable X

Let a continuous random variable X have a probability density function $f(x)$ defined over the interval $a \le x \le b$. Then the expected value of X is given by

$$E(X) = \mu = \int_a^b x\,f(x)\,dx$$

In some cases, the integral is written as

$$E(X) = \mu = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

where $-\infty$ represents the lowest possible value of x and ∞ represents the highest possible value of x.

8.2

WORKED EXAMPLE 7 Finding the expected value of a continuous random variable with one component

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} 4x^3 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value of X.

Steps	Working
1 Write the integral for the formula $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$ and simplify.	$E(x) = \int_0^1 x \times 4x^3\, dx$ $= \int_0^1 4x^4\, dx$
2 Evaluate the integral.	$E(X) = \left[\frac{4x^5}{5}\right]_0^1$ $= \frac{4(1)^5}{5} - 0$ $= \frac{4}{5}$

Exam hack

When a probability density function $f(x)$ has a piece-wise definition with two or more components, you will be required to write two or more separate integrals to calculate the value of $E(X)$.

WORKED EXAMPLE 8 Finding the expected value of a continuous random variable with two components

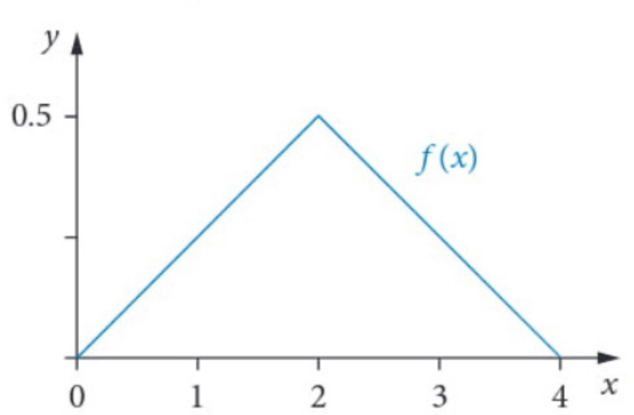
The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{4} & 0 \le x < 2 \\ 1 - \frac{x}{4} & 2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(X)$.

Steps	Working
1 Write the two integrals for the formula $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$ and simplify.	$E(X) = \int_0^2 x \times \frac{x}{4}\, dx + \int_2^4 x \times \left(1 - \frac{x}{4}\right) dx$ $= \int_0^2 \frac{x^2}{4}\, dx + \int_2^4 \left(x - \frac{x^2}{4}\right) dx$
2 Evaluate the integrals.	$E(X) = \left[\frac{x^3}{12}\right]_0^2 + \left[\frac{x^2}{2} - \frac{x^3}{12}\right]_2^4$ $= \left[\frac{2^3}{12} - 0\right] + \left[\left(\frac{4^2}{2} - \frac{4^3}{12}\right) - \left(\frac{2^2}{2} - \frac{2^3}{12}\right)\right]$ $= \frac{8}{12} + \left[\left(8 - \frac{64}{12}\right) - \left(2 - \frac{8}{12}\right)\right]$ $= 2$

The graph of the pdf shows by symmetry that the mean is 2.



In some cases, $xf(x)$ will produce a function that we do not have the skills and techniques required to integrate by hand. This is where CAS can be useful.

USING CAS 3 Calculating the expected value

A continuous random variable X is defined by the following probability density function.

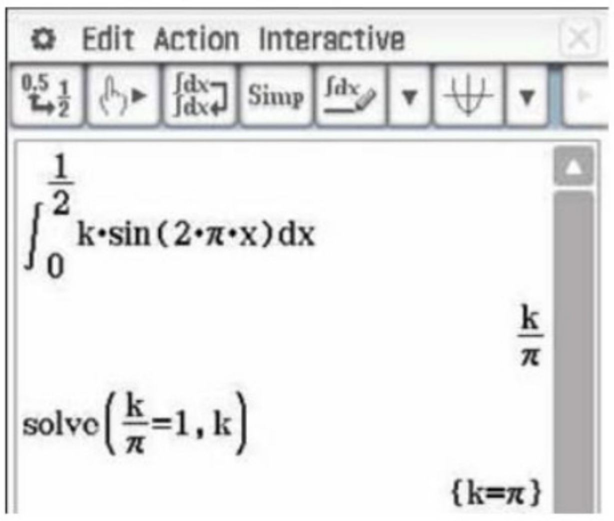
$$f(x) = \begin{cases} k\sin(2\pi x) & 0 \le x \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

a Show that the value of k is π.

b Determine $\text{E}(X)$.

ClassPad

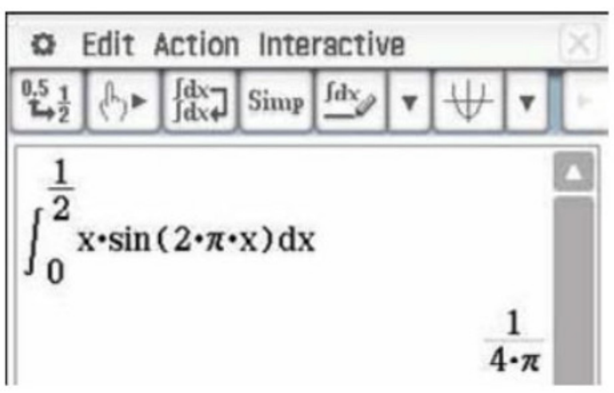
a



1 Establish the definite integral giving a total area under the curve.

2 Solve the result equal to 1.

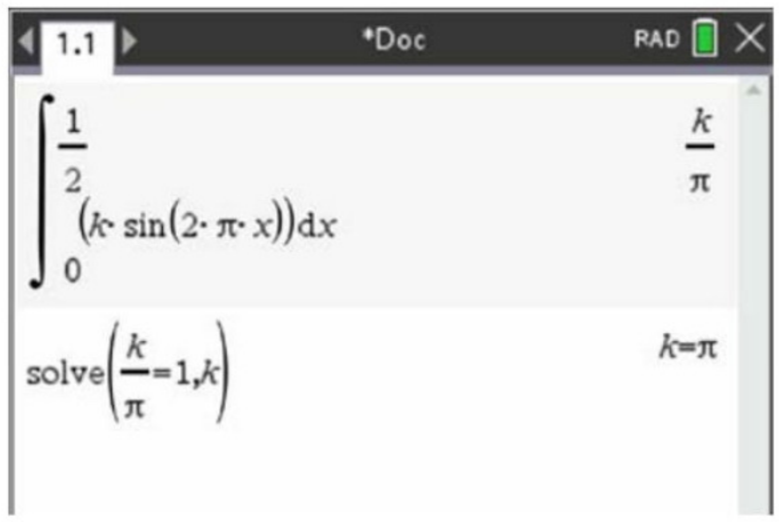
b



Establish and evaluate the definite integral for the expected value using $\text{E}(X) = \int_{-\infty}^{\infty} xf(x)\,dx$.

TI-Nspire

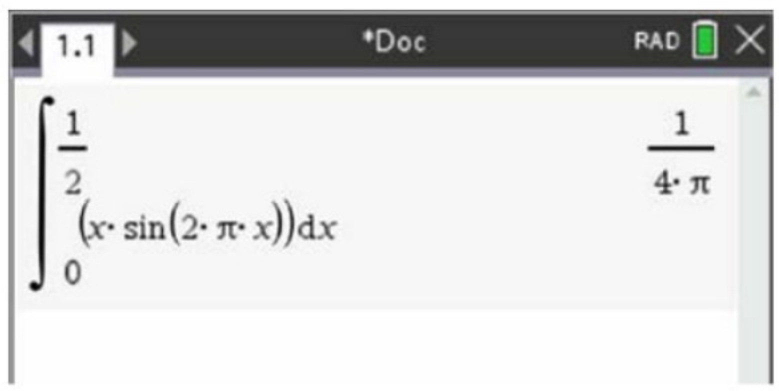
a



1 Establish the definite integral giving a total area under the curve.

2 Solve the result equal to 1.

b



Establish and evaluate the definite integral for the expected value using $\text{E}(X) = \int_{-\infty}^{\infty} xf(x)\,dx$.

This gives the expected value of $\frac{1}{4\pi}$.

The median and other percentiles

In addition to the expected value (or mean) being used to measure the centre of a continuous random variable, we can also encounter problems that involve calculating the **median** of a continuous random variable. Recall that the **median** is the middle score of a random variable, when the outcomes are in ascending or descending order.

The median of a continuous random variable X

For a continuous random variable X, let $x = m$ be the median. Then

$$\text{P}(X < m) = \text{P}(X \le m) = \text{P}(X > m) = \text{P}(X \ge m) = 0.5$$

9780170477536

The value of m can be solved for using an appropriate definite integral of the probability density function $f(x)$ defined over the interval $a \leq x \leq b$.

$$\int_a^m f(x)\,dx = \int_m^b f(x)\,dx = 0.5$$

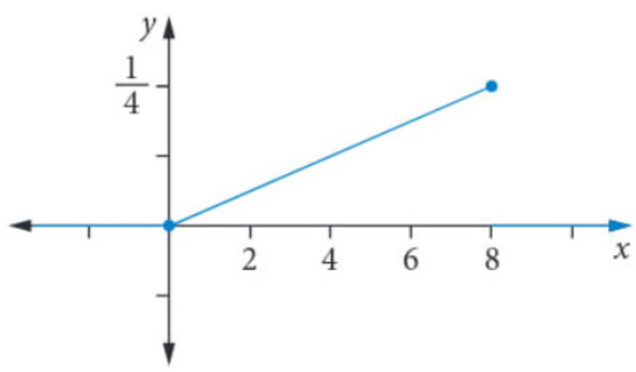
8.2

WORKED EXAMPLE 9 **Finding the median of a continuous random variable**

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \dfrac{x}{32} & 0 \leq x \leq 8 \\ 0 & \text{otherwise} \end{cases}$$

Find the median, m, of X.

Steps	Working
1 Sketch the probability density function.	
2 Establish and solve the integral equation of the 50th percentile.	$\int_0^m \frac{x}{32}\,dx = 0.50$ $\left[\frac{x^2}{64}\right]_0^m = \frac{1}{2}$ $\frac{m^2}{64} = \frac{1}{2}$ $m^2 = 32$ $m = \pm\sqrt{32} = \pm 4\sqrt{2}$ $\therefore m = 4\sqrt{2}$ (as $m > 0$)

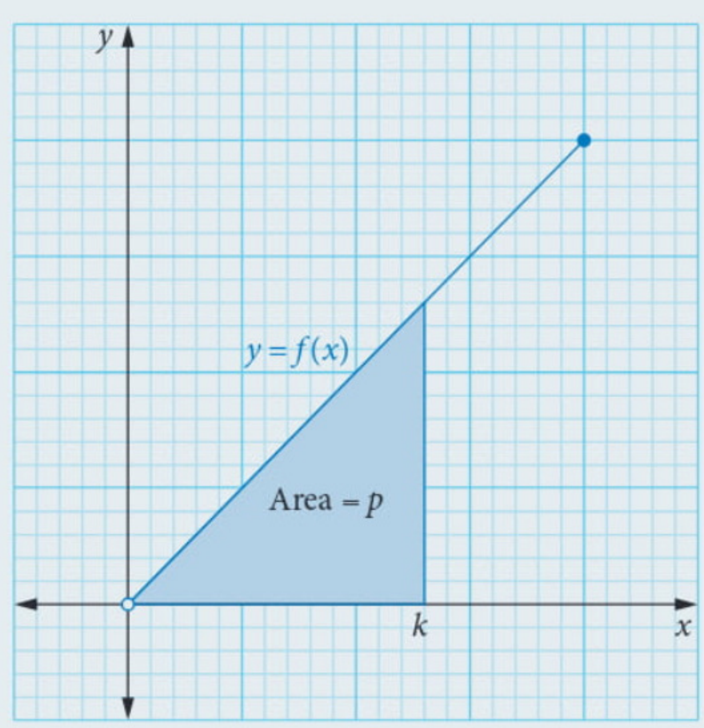
The median is a specific example of a **percentile** of a continuous random variable X, which is a value of x for which a certain percentage of scores, p%, fall below that value. That is, the median is the 50th percentile. Other common percentiles include

- the 25th percentile or lower quartile
- the 75th percentile or upper quartile.

A percentile of a continuous random variable X

For a continuous random variable X with a probability density function $f(x)$, the value of the pth percentile ($x = k$) is determined by the integral equation

$$P(X \leq k) = \frac{p}{100} \Rightarrow \int_{-\infty}^{k} f(x)\,dx = \frac{p}{100}.$$

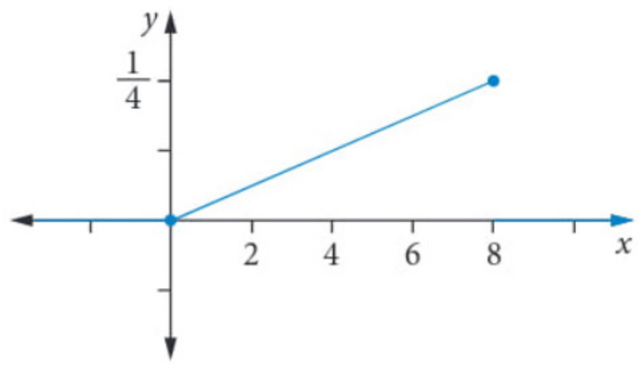


WORKED EXAMPLE 10 Finding a percentile of a continuous random variable

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{32} & 0 \leq x \leq 8 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k such that $P(X < k) = 0.75$.

Steps	Working
1 Sketch the probability density function.	
2 Establish and solve the integral equation of the 75th percentile.	$\int_0^k \frac{x}{32}\,dx = 0.75$ $\left[\frac{x^2}{64}\right]_0^k = \frac{3}{4}$ $\frac{k^2}{64} - 0 = \frac{3}{4}$ $k^2 = 48$ $k = \pm\sqrt{48}$ $= \pm 4\sqrt{3}$ $\therefore\ k = 4\sqrt{3} \quad (\text{as } k > 0)$

Variance and standard deviation as measures of spread

Similarly, recall from Chapter 5 that the variance and standard deviation were useful measures of spread of a random variable X about its mean, whereby

- the variance is the weighted average of the squared deviations from the mean
- the standard deviation is the square root of the variance.

Once again, in discrete cases we used the summation notation below as the values of $p(x)$ were defined for discrete values of x.

$$\text{Var}(x) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

However, now that X is continuous, $p(x)$ is also a continuous curve and so the summation again takes the form of a definite integral.

The variance and standard deviation of a continuous random variable X

Let a continuous random variable X have a probability density function $f(x)$ defined over the interval $a \le x \le b$. Then the variance of X is given by

$$\text{Var}(X) = \sigma^2 = \int_a^b (x-\mu)^2 f(x)\,dx.$$

In some cases, the integral is written as

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

where $-\infty$ represents the lowest possible value of x and ∞ represents the highest possible value of x.

Remember, the more efficient computational formula for variance can also be used:

$$\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2 = \int_a^b x^2 f(x)\,dx - \mu^2$$

The standard deviation is then $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

8.2

WORKED EXAMPLE 11 **Finding the variance and standard deviation of a continuous random variable**

The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} 4x^3 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the variance and standard deviation of X.

Steps	Working
1 Write the formula for the variance and calculate E(X) and E(X^2) using $\text{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$. Note: μ or E(X) was calculated in Worked example 7.	$\text{Var}(X) = \text{E}(X^2) - \mu^2$ $\text{E}(X) = \int_0^1 x \times 4x^3\,dx = \frac{4}{5}$ $\text{E}(X^2) = \int_0^1 x^2 \times 4x^3\,dx = \int_0^1 4x^5\,dx$ $\text{E}(X^2) = \left[\frac{4x^6}{6}\right]_0^1 = \frac{2(1^6)}{3} - 0 = \frac{2}{3}$
2 Substitute into the variance formula.	$\text{Var}(X) = \text{E}(X^2) - \mu^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{3} - \frac{16}{25} = \frac{2}{75}$
3 Use the fact $\text{SD}(X) = \sqrt{\text{Var}(X)}$.	$\text{SD}(X) = \sqrt{\frac{2}{75}} = \frac{\sqrt{6}}{15}$

Linear changes of scale and origin

Recall the ideas of a linear change of scale and origin from Chapter 5, such that

- a change of scale of a random variable X multiplies all values of x by a scalar multiple a, i.e. $X \to aX$
- a change of origin of a random variable X adds a value b (or subtracts b from) all values of x, i.e. $X \to X \pm b$.

The effects of these changes of scale and origin have the same effect on the mean, variance and standard deviation of a continuous random variable as they did to a discrete random variable due to the properties of definite integration.

Expected value, variance and standard deviation of $Y = aX + b$

For a linear transformation $Y = aX + b$ of a continuous random variable X:

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$$

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

$$\mathrm{SD}(aX + b) = |a|\,\mathrm{SD}(X)$$

WORKED EXAMPLE 12 **Finding the expected value, variance and standard deviation of $aX + b$**

A continuous random variable X has a mean of 4 and variance of 16. A continuous random variable Y is defined such that $Y = -\frac{1}{2}X + 5$. Determine

a $\mathrm{E}(Y)$

b $\mathrm{Var}(Y)$

c $\mathrm{SD}(Y)$.

Steps	Working
a Apply the formula $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$.	$\mathrm{E}(Y) = \mathrm{E}\left(-\frac{1}{2}X + 5\right)$ $= -\frac{1}{2}(4) + 5$ $= 3$
b Apply the formula $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.	$\mathrm{Var}(Y) = \mathrm{Var}\left(-\frac{1}{2}X + 5\right)$ $= \frac{1}{4}(16)$ $= 4$
c Apply the formula $\mathrm{SD}(aX + b) = |a|\,\mathrm{SD}(X)$.	$\mathrm{SD}(Y) = \frac{1}{2}(4)$ $= 2$ or $\sqrt{\mathrm{Var}(Y)} = \sqrt{4}$ $= 2$

EXERCISE 8.2 Measures of centre and spread

ANSWERS p. 404

8.2

Recap

1 The probability density function for the continuous random variable X is given by

$$f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The value of $\mathrm{P}\left(X > \frac{1}{2}\right)$ is

A $\frac{1}{27}$ **B** $\frac{1}{8}$ **C** $\frac{2}{3}$ **D** $\frac{7}{8}$ **E** $\frac{26}{27}$

2 If X is a continuous random variable such that $\mathrm{P}(X > 5) = a$ and $\mathrm{P}(X > 8) = b$ then $\mathrm{P}(X < 5 \mid X < 8)$ is

A $\frac{a}{b}$ **B** $\frac{a-b}{1-b}$ **C** $\frac{1-b}{1-a}$ **D** $\frac{ab}{1-b}$ **E** $\frac{a-1}{b-1}$

Mastery

3 WORKED EXAMPLE 7 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{1}{288}(12x - x^2) & 0 \le x \le 12 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value of X.

4 WORKED EXAMPLE 8 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{3x}{8} & 0 \le x \le 2 \\ 3 - \frac{9}{8}x & 2 < x \le \frac{8}{3} \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathrm{E}(X)$.

5 Using CAS 3 A continuous random variable X is defined by the following probability density function.

$$f(x) = \begin{cases} k\sin(\pi x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Show that the value of k is $\frac{\pi}{2}$.

b Determine $\mathrm{E}(X)$.

6 WORKED EXAMPLE 9 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{12} & 1 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

Find the median, m, of X.

7 WORKED EXAMPLE 10 The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{8} & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of k such that $P(X < k) = 0.3$.

8 WORKED EXAMPLE 11 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the variance and standard deviation of X.

9 WORKED EXAMPLE 12 A continuous random variable X has a mean of 3 and variance of 9. A continuous random variable Y is defined such that $Y = \frac{2}{3}X - 1$. Determine

a $E(Y)$ **b** $Var(Y)$ **c** $SD(Y)$.

Calculator-free

10 (6 marks) The continuous random variable X, with probability density function $p(x)$ defined over $a \le x \le b$, has mean 2 and variance 5.

a Determine the value of $\int_a^b x^2 p(x)\,dx$. (2 marks)

b Determine $E(3X + 2)$. (2 marks)

c Determine $Var\left(-\frac{1}{5}X - 1\right)$. (2 marks)

11 (4 marks) A continuous random variable X has the probability density function f given by

$$f(x) = \begin{cases} \frac{k}{x^2} & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = 2$. (2 marks)

b Determine $E(X)$. (2 marks)

12 (3 marks) A continuous random variable, X, has a probability density function given by

$$f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

The median of X is m such that $P(X < m) = 0.5$. Determine the exact value of m.

13 (7 marks) The probability density function for a continuous random variable X is given by

$$f(x) = \begin{cases} ax(5 - x) & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

where a is a positive constant.

a Show that $a = \frac{6}{125}$. (3 marks)

b Explain why $E(X) = \frac{5}{2}$. (1 marks)

c Show that $SD(X) = \frac{\sqrt{5}}{2}$. (3 marks)

Calculator-assumed

14 © SCSA MM2018 Q10c (2 marks) The following function is a probability density function on the given interval

$$f(x) = \begin{cases} ax^2(x-2) & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the median of the distribution.

15 © SCSA MM2017 Q11bc (7 marks) A pizza shop estimates that the time X hours to deliver a pizza from when it is ordered is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \frac{4}{3} - \frac{2}{3}x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

a Calculate the mean delivery time to the nearest minute. (3 marks)

b Calculate the standard deviation of the delivery time to the nearest minute. (4 marks)

16 (10 marks) A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{1}{8}e^{-\frac{x}{8}} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

a Show that $f(x)$ is a valid probability density function for a continuous random variable. (3 marks)

b Determine

i $\text{E}(X)$ (2 marks)

ii $\text{Var}(X)$. (3 marks)

c Find the value of k, correct to three decimal places, for which $\text{P}(X < k) = 0.25$. (2 marks)

17 (10 marks) For the continuous random variable X with probability density function

$$f(x) = \begin{cases} \log_e(x) & 1 \le x \le e \\ 0 & \text{otherwise} \end{cases}$$

determine the following correct to four decimal places, where appropriate.

a $\text{P}(X > 1 \mid X < 2)$ (2 marks)

b k such that $\text{P}(X < k) = 0.8$ (2 marks)

c $\text{E}(X)$ (2 marks)

d $\text{Var}(X)$ (2 marks)

e $\text{SD}(Y)$ where $Y = -2X$ (2 marks)

18 (5 marks) FullyFit is an international company that owns and operates many fitness centres (gyms) in several countries. At every one of FullyFit's gyms, each member agrees to have his or her fitness assessed every month by undertaking a set of exercises called S. There is a five-minute time limit on any attempt to complete S and if someone completes S in less than three minutes, they are considered fit.

When FullyFit surveyed all its gyms throughout the world, it was found that the time taken by members to complete S is a continuous random variable X, with a probability density function g, as defined below.

$$g(x) = \begin{cases} \dfrac{(x-3)^3 + 64}{256} & 1 \le x \le 3 \\ \dfrac{x+29}{128} & 3 < x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

a Find $E(X)$, correct to four decimal places. (2 marks)

b In a random sample of 200 FullyFit members, how many members would be expected to take more than four minutes to complete S? Give your answer to the nearest integer. (3 marks)

19 (8 marks) Rebecca's Robotics manufactures three types of components for robots: sensors, motors and controllers.

The weight, w, in grams, of controllers is modelled by the following probability density function.

$$C(w) = \begin{cases} \dfrac{3}{640000}(330-w)^2(w-290) & 290 \le w \le 330 \\ 0 & \text{otherwise} \end{cases}$$

a Determine the mean weight, in grams, of the controllers. (2 marks)

b Determine the probability that a randomly selected controller weighs less than the mean weight of the controllers. (2 marks)

c Determine the standard deviation of the weight, in grams, of the controllers. (2 marks)

d Determine the probability that a randomly selected controller weighs more than one standard deviation greater than the mean weight of the controllers. (2 marks)

8.3 Uniform and triangular distributions

Uniform continuous random variables

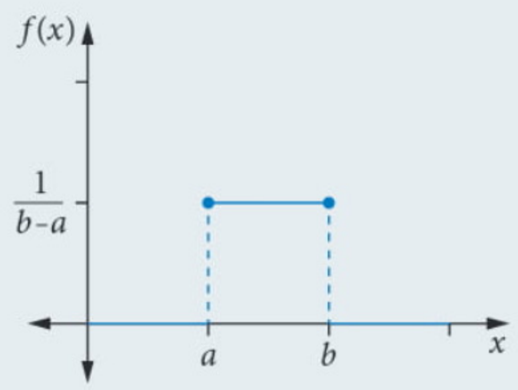
In Chapter 5 we examined simple examples of discrete random variables that had the same probability for each outcome $x \in X$. Suppose that is now the case for a continuous random variable X defined over the interval $a \leq x \leq b$ such that each value of x has an equally likely chance of occurring. This random variable can be described as a **uniform continuous random variable**.

Continuous uniform distribution

For a continuous random variable X, if X is uniformly distributed over the interval $a \leq x \leq b$, then the probability density function of X is defined as

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

A uniform continuous random variable can be denoted using $X \sim U[a, b]$.



The probability density function creates a rectangular-shaped graph.

Video playlist
Uniform continuous random variables

Exam hack

In most cases, probabilities from a continuous uniform distribution can be found simply using the area of rectangles rather than definite integration! Regardless, always be sure to show the appropriate working, whether it is the area of the rectangles used or the definite integrals used.

WORKED EXAMPLE 13 Calculating probabilities from a continuous uniform distribution

A uniformly distributed continuous random variable X has the probability density function

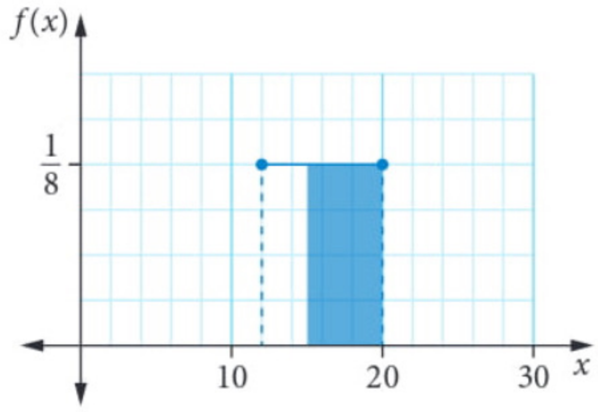
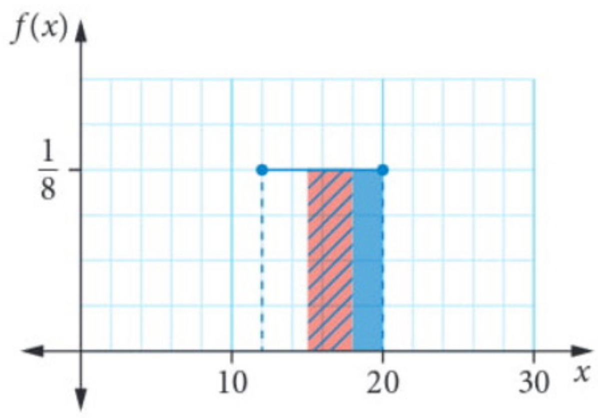
$$f(x) = \begin{cases} \frac{1}{8} & 12 \leq x \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

Determine

a $P(X = 13)$

b $P(X > 15)$

c $P(X \leq 18 \mid X > 15)$

Steps	Working
a Recognise that $P(X = k) = 0$ for a continuous random variable.	$P(X = 13) = 0$
b **1** Draw a sketch of $f(x)$ and shade the appropriate region under the curve.	
2 Calculate the area of the rectangle.	$P(X > 15) = P(15 < X \le 20)$ $= 5\left(\frac{1}{8}\right)$ $= \frac{5}{8}$
c **1** Establish the conditional probability formula $P(A\|B) = \frac{P(A \cap B)}{P(B)}$.	$(X \le 18 \| X > 15) = \frac{P(15 < X \le 18)}{P(X > 15)}$
2 Represent the probability diagrammatically.	
3 Calculate the areas of the rectangles and, hence, the probability.	$= \frac{\left(\frac{3}{8}\right)}{\left(\frac{5}{8}\right)}$ $= \frac{3}{5}$

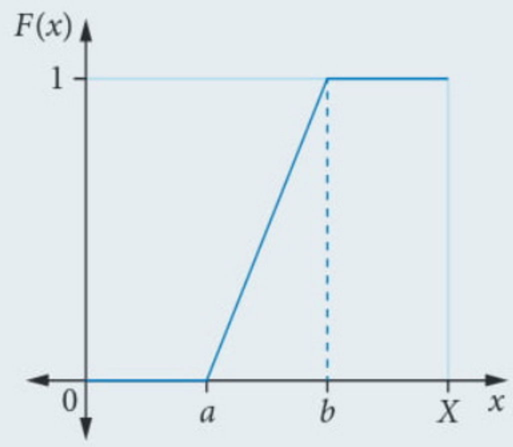
The cumulative distribution function of a uniform continuous random variable can also be obtained using integration, such that for $a \le x \le b$,

$$\begin{aligned} F(x) &= \int_a^x \frac{1}{b-a}\,dt \\ &= \left[\frac{t}{b-a}\right]_a^x \\ &= \frac{x}{b-a} - \frac{a}{b-a} \\ &= \frac{x-a}{b-a} \end{aligned}$$

Cumulative distribution function of a continuous uniform distribution

For a uniformly distributed continuous random variable $X \sim U[a,b]$, the cumulative distribution function of X is defined as

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



8.3

WORKED EXAMPLE 14 **Determining the probability density function from a cumulative distribution function**

A uniformly distributed continuous random variable Y has a cumulative distribution function given by

$$F(y) = \begin{cases} 0 & y < a \\ \dfrac{y-a}{10} & a \le y \le 15 \\ 1 & y > 15 \end{cases}$$

a Determine the value of a.

b State the probability density function of Y.

c Hence, or otherwise, calculate $P(Y < 12 \mid Y < 14)$.

Steps	Working
a Recognise that $b - a = 10$ and solve for a.	$15 - a = 10$ $a = 5$
b Establish the piece-wise-defined function for the uniform continuous random variable either using $f(x) = \dfrac{1}{b-a}$ or $f(x) = \dfrac{d}{dx}(F(x))$.	$f(y) = \begin{cases} \dfrac{1}{10} & 5 \le y \le 15 \\ 0 & \text{otherwise} \end{cases}$
c 1 Recognise that in $P(A \mid B)$, set A is contained within set B and so $P(A \mid B) = \dfrac{P(A)}{P(B)}$.	$P(Y < 12 \mid Y < 14) = \dfrac{P(Y < 12)}{P(Y < 14)}$
2 Use the cumulative distribution function to evaluate $P(Y < 12)$ and $P(Y < 14)$, or consider the area of rectangles.	$= \dfrac{\left(\dfrac{12-5}{10}\right)}{\left(\dfrac{14-5}{10}\right)}$ $= \dfrac{7}{9}$

When considering the expected value of a random variable with a symmetrical distribution such as the continuous uniform distribution, note that the symmetry of the distribution means that the expected value should lie in the very centre of the interval of x values and, hence, can be calculated using the average of the values of a and b.

Expected value of a uniform continuous random variable

For a uniformly distributed continuous random variable, $X \sim U[a, b]$,

$$\mathrm{E}(X) = \frac{a+b}{2}$$

This can also be proven using integration, i.e. $\mathrm{E}(X) = \int_a^b x\left(\frac{1}{b-a}\right)dx$.

Note that due to the symmetry of the distribution, the median of a uniform continuous random variable is the same as the mean.

We can then use integration and the computational formula for variance, $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$ to derive a formula for the variance and standard deviation of a uniformly distributed continuous random variable.

$$\begin{aligned}\mathrm{E}(X^2) &= \int_a^b x^2\left(\frac{1}{b-a}\right)dx \\ &= \frac{1}{b-a}\int_a^b x^2\,dx \\ &= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\ &= \frac{1}{b-a} \times \frac{b^3-a^3}{3}\end{aligned}$$

Using the factorisation $b^3 - a^3 = (b-a)(b^2 + ab + a^2)$,

$$\mathrm{E}(X^2) = \frac{b^2+ab+a^2}{3}$$

$$\begin{aligned}\mathrm{Var}(X) &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} \\ &= \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12} \\ &= \frac{b^2-2ab+a^2}{12}\end{aligned}$$

Using the factorisation $b^2 - 2ab + a^2 = (b-a)^2$,

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Variance and standard deviation of a uniform continuous random variable

For a uniformly distributed continuous random variable, $X \sim U[a, b]$,

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

and

$$\text{SD}(X) = \frac{b-a}{\sqrt{12}}.$$

8.3

Exam hack

The formulas for the expected value, variance and standard deviation of uniform continuous random variable are not on the formula sheet. Write them on your notes, but be sure to pay attention to the number of marks in a question to calculate mean, variance or standard deviation (more than 2 marks per statistic) or any specific instructions to *use integration*. In these cases, the formulas may not be helpful!

WORKED EXAMPLE 15 **Finding the expected value, variance and standard deviation of a uniform continuous random variable**

A uniformly distributed continuous random variable X is defined such that $X \sim U[-10, 10]$.

Find the

a probability density function of X

b expected value of X

c variance of X

d standard deviation of X.

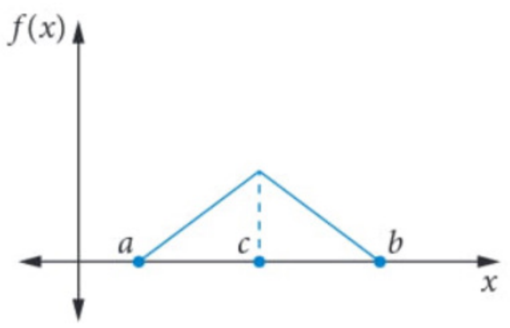
Steps	Working
a **1** Identify the values of a and b.	$a = -10$, $b = 10$
2 Use the piece-wise definition of $f(x)$. $f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$f(x) = \begin{cases} \frac{1}{20} & -10 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$
b Use the formula $\text{E}(X) = \frac{a+b}{2}$.	$\text{E}(X) = \frac{-10+10}{2} = 0$
c Use the formula $\text{Var}(X) = \frac{(b-a)^2}{12}$.	$\text{Var}(X) = \frac{20^2}{12} = \frac{400}{12} = \frac{100}{3}$
d Use the formula $\text{SD}(X) = \sqrt{\text{Var}(X)}$.	$\text{SD}(X) = \sqrt{\frac{100}{3}} = \frac{10\sqrt{3}}{3}$

Triangular continuous random variables

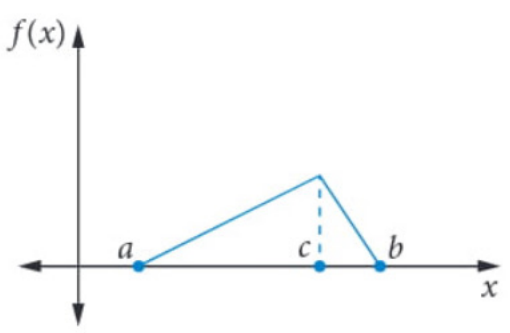
In some previous examples, such as Worked example 8, we were exposed to a particular type of probability density function for a special continuous random variable. You may recall seeing a probability density function with a graph in the shape of a triangle. Continuous random variables with triangular-shaped graphs are called **triangular continuous random variables**.

Triangular distributions can be both symmetrical and asymmetrical. Suppose a continuous random variable X is triangularly distributed with a probability density function $f(x)$ defined over the interval $a \le x \le b$. Let the 'peak' of the triangle occur at $x = c$. This peak can be considered the **mode** of the triangular distribution as it is the value of x with the highest value of $f(x)$.

If symmetrical, then the mode occurs halfway between a and b and, hence, is the mean and median of the distribution, i.e. $c = \mathrm{E}(X) = \dfrac{a+b}{2}$.



If asymmetrical, then the mean will **not** be the same as the mode.



Maximum value of $f(x)$ in a triangular distribution

Given that the area of the triangle with base $(b - a)$ and a perpendicular height $h = f(c)$ needs to equal 1, then

$$\frac{1}{2}(b-a)f(c) = 1$$

$$f(c) = \frac{2}{b-a}$$

For a triangular continuous random variable X, then the maximum value of $f(x)$ at the mode $x = c$ is

$$f(c) = \frac{2}{b-a}$$

Although the probability density function, expected value and variance of a triangular distribution can be generalised regardless of the symmetry, it is not an explicit part of this course. As a result, it is often more useful to consider the function as a piece-wise-defined function with linear components, as we saw in previous examples, and use the appropriate geometric and integration techniques to solve any related problems.

WORKED EXAMPLE 16 Using a symmetrical triangular distribution

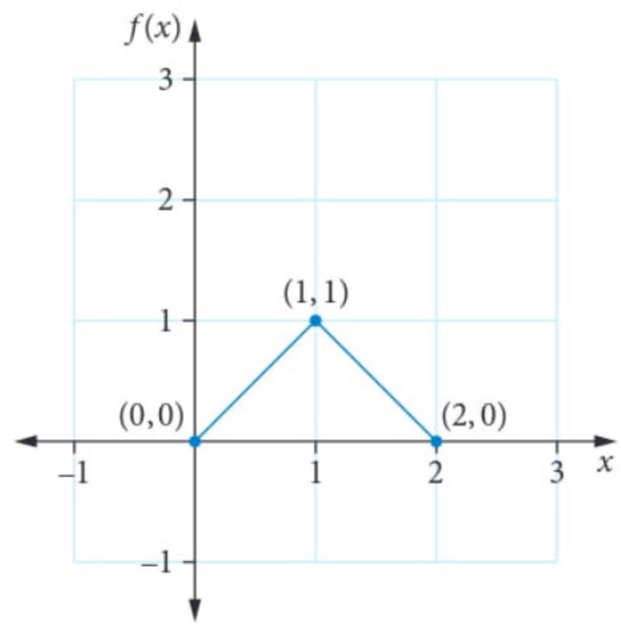
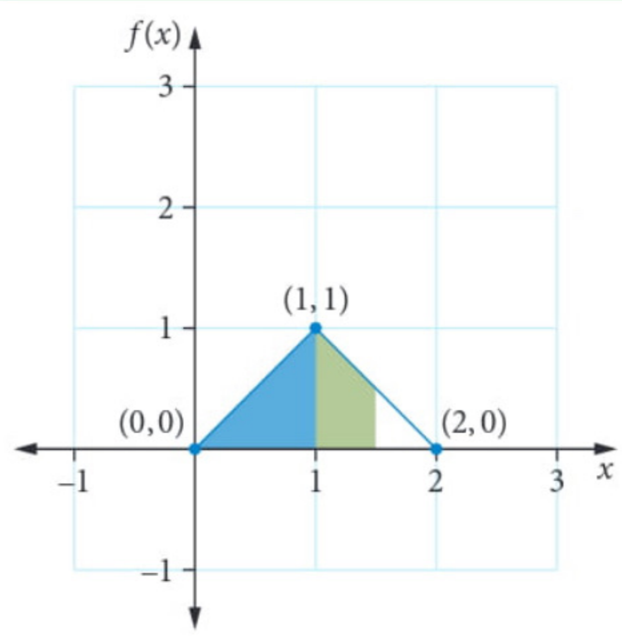
Consider the following probability density function for the continuous random variable X.

$$f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2 - x & 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(x)$ labelling the coordinates of all critical features. Hence, describe the shape of the distribution.

b Determine

i $P(X < 1.5)$ **ii** $P(X < 2 | X > 1.5)$ **iii** $E(X)$ **iv** $\text{Var}(X)$.

Steps	Working
a 1 Sketch each linear component over the given domains.	
2 Label all axes intercepts and the 'peak'.	
3 Describe the shape of the distribution, including the parameters.	It is a symmetrical triangular distribution over the interval $0 \le x \le 2$.
b i 1 Identify the region on the graph.	
2 Use a geometric approach (i.e. triangle and trapezium) to calculate the probability as the area under the curve.	$P(X < 1.5) = \frac{1}{2} + \frac{1}{2}\left(1 + \frac{1}{2}\right)\left(\frac{1}{2}\right)$ $= \frac{1}{2} + \frac{1}{4}\left(\frac{3}{2}\right)$ $= \frac{7}{8}$

ii **1** Use the conditional probability formula $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.	$P(X < 2 \mid X > 1.5) = \dfrac{P(1.5 < X < 2)}{P(X > 1.5)}$
2 Use the complement to find $P(X > 1.5)$ given part **a**.	$= \dfrac{\frac{1}{2}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{1 - \frac{7}{8}}$ $= 1$
	Exam hack Be sure to look out for 'trick' questions like this one. Of course it is certain that $x < 2$ if $x > 1.5$, as the upper bound is 2.
iii Use the symmetry to identify the mean.	$E(X) = 1$
iv **1** Find $E(X^2)$ using integration.	$E(X^2) = \int_0^1 x^2(x)\,dx + \int_1^2 x^2(2-x)\,dx$ $= \int_0^1 x^3\,dx + \int_1^2 2x^2 - x^3 dx$
2 Simplify the result for ease of computation.	$= \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2$ $= \frac{1}{4} + \left(\frac{16}{3} - \frac{16}{4} - \frac{2}{3} + \frac{1}{4}\right)$ $= -\frac{14}{4} + \frac{14}{3}$ $= \frac{-42 + 56}{12}$ $= \frac{14}{12}$ $= \frac{7}{6}$
3 Use $\text{Var}(X) = E(X^2) - E(X)^2$.	$\text{Var}(X) = \frac{7}{6} - 1^2$ $= \frac{1}{6}$

WORKED EXAMPLE 17 Using an asymmetrical triangular distribution

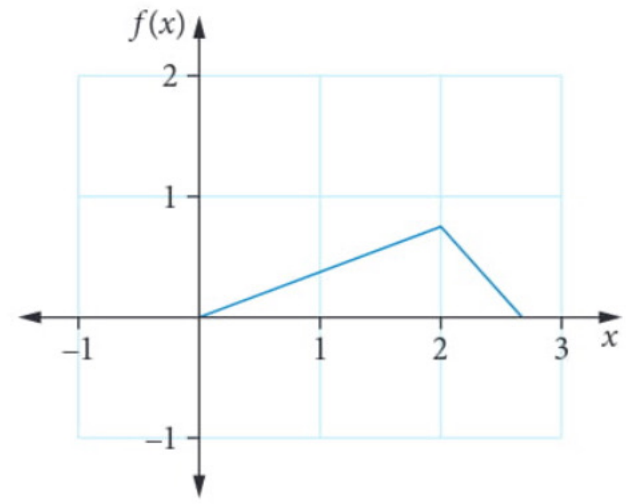
Consider the following probability density function for the continuous random variable X.

$$f(x) = \begin{cases} \dfrac{3x}{8} & 0 \le x \le 2 \\ 3 - \dfrac{9}{8}x & 2 < x \le \dfrac{8}{3} \\ 0 & \text{otherwise} \end{cases}$$

a Show that it is a valid probability density function for X.

b Determine the 30th percentile.

c Show the use of integration to determine $E(X)$.

Steps	Working
a **1** Identify the two conditions of a valid pdf. • $f(x) \geq 0$ for all x in $a \leq x \leq b$ • $\int_a^b f(x)\,dx = 1$	The end points of the line segment with equation $\frac{3x}{8}$ are $(0, 0)$ and $\left(2, \frac{3}{4}\right)$. The end points of the line segment with equation $3 - \frac{9}{8}x$ are $\left(2, \frac{3}{4}\right)$ and $\left(\frac{8}{3}, 0\right)$.
2 Show the first condition graphically or numerically.	
3 Show the second condition geometrically using the area of triangles or algebraically using integration.	And so $f(x) \geq 0$ for all values of x in the domain $0 \leq x \leq \frac{8}{3}$. $\int_0^{\frac{8}{3}} f(x)\,dx = \frac{1}{2}(2)\left(\frac{3}{4}\right) + \frac{1}{2}\left(\frac{2}{3}\right)\left(\frac{3}{4}\right)$ $= \frac{3}{4} + \frac{1}{4}$ $= 1$ Therefore, $f(x)$ is a valid pdf.
b **1** Recognise that $P(X < 2) = 0.75$ and so the 30th percentile lies in the first triangular region. Assign a variable to the value of the percentile.	Let the 30th percentile be $x = k$.
2 Establish the appropriate integral (or area formula of a triangle) to calculate the 30th percentile.	$\int_0^k \frac{3x}{8}\,dx = \frac{3}{10}$ $\left[\frac{3x^2}{16}\right]_0^k = \frac{3}{10}$ $\frac{3k^2}{16} = \frac{3}{10}$ Alternatively, $A = \frac{1}{2}bh$ $\frac{3}{10} = \frac{1}{2}k\left(\frac{3k}{8}\right)$ $\frac{3k^2}{16} = \frac{3}{10}$ $k^2 = \frac{8}{5}$ $k = \pm\frac{2\sqrt{10}}{5}$ $k = \frac{2\sqrt{10}}{5}$ (as $k > 0$)

c 1 Establish the sum of two definite integrals for E(X).

$$\text{E}(X) = \int_0^2 x\left(\frac{3x}{8}\right)dx + \int_2^{\frac{8}{3}} x\left(3 - \frac{9x}{8}\right)dx$$

$$= \int_0^2 \frac{3x^2}{8}dx + \int_2^{\frac{8}{3}} 3x - \frac{9}{8}x^2 dx$$

2 Evaluate the definite integrals.

$$= \left[\frac{x^3}{8}\right]_0^2 + \left[\frac{3x^2}{2} - \frac{9}{24}x^3\right]_2^{\frac{8}{3}}$$

$$= 1 + \left(\frac{3}{2}\left(\frac{64}{9}\right) - \frac{9}{24}\left(\frac{512}{27}\right) - 6 + 3\right)$$

$$= 1 + \frac{32}{3} - \frac{64}{9} - 3$$

$$= \frac{14}{9}$$

Problems involving the uniform, triangular and binomial distributions

In some practical problems where continuous random variables are modelled by uniform or triangular distributions, we may be asked to solve a binomial problem for which the probability of success is to be determined from the uniform or triangular distribution. The features of a binomially distributed random variable were explored in Chapter 5.

Features of a binomial distribution

For $X \sim \text{Bin}(n, p)$, then

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

$$\text{E}(X) = np$$

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

$$\text{SD}(X) = \sqrt{np(1 - p)}$$

In these problem types, be sure to define a new random variable when you notice the context change to a binomial situation.

WORKED EXAMPLE 18 **Modelling using the uniform and binomial distributions**

The local supermarket packages potatoes in bags with a labelled weight of 4 kg. The weight of the bags of potatoes is found to be uniformly distributed, with weights ranging from 3980 g to 4040 g. Let the weight of a bag of potatoes be represented by the continuous random variable W.

a Determine the mean and standard deviation of W.

b Determine the probability that a randomly selected bag of potatoes weighs less than the labelled weight of 4 kg.

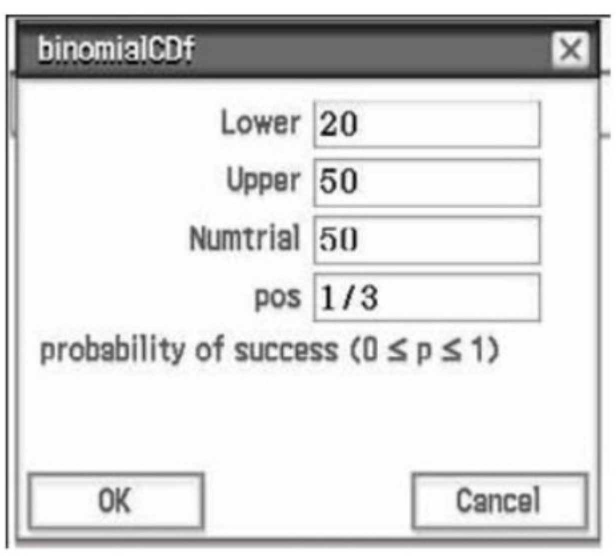
Suppose 50 different bags of potatoes are to be sampled.

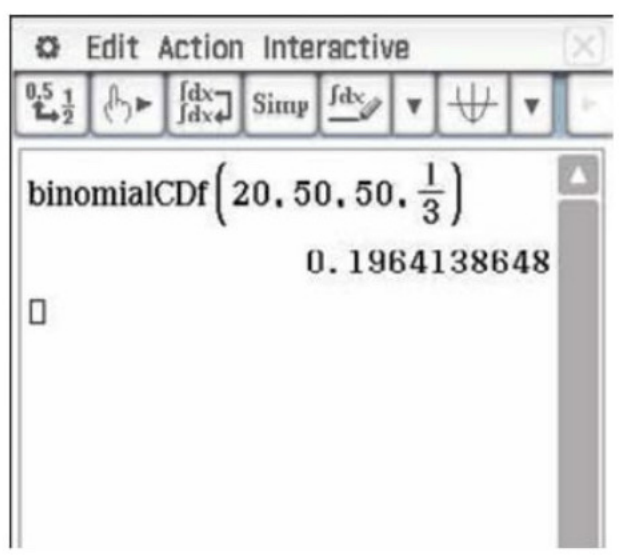
c How many of the 50 bags are expected to weigh less than 4 kg? Answer to the nearest whole bag.

d Determine the probability, correct to four decimal places, that at least 20 of the 50 bags will have a weight less than 4 kg.

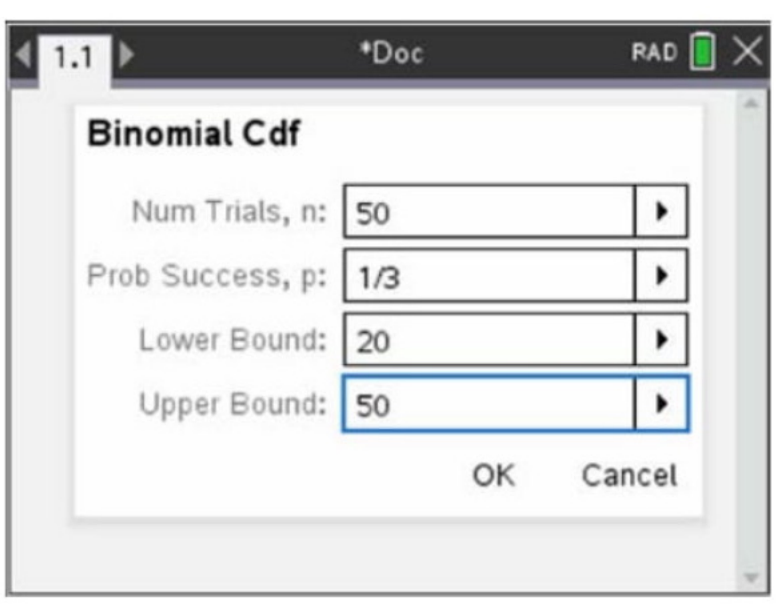
Steps	Working
a Use the formulas $E(X) = \frac{a+b}{2}$ and $SD(X) = \frac{b-a}{\sqrt{12}}$.	$E(W) = \frac{3980 + 4040}{2}$ $= \frac{8020}{2}$ $= 4010\text{ g}$ $SD(W) = \frac{4040 - 3980}{\sqrt{12}} = \frac{60}{\sqrt{12}}$ $= 10\sqrt{3}$ or 17.32 g
b **1** Determine the value of $f(x)$ using $\frac{1}{b-a}$.	$f(w) = \frac{1}{4040 - 3980} = \frac{1}{60}$
2 Consider the area of the rectangle as the probability.	$P(W < 4000) = 20\left(\frac{1}{60}\right) = \frac{1}{3}$
c **1** Define a new binomial random variable.	Let X be the number of bags out of 50 with a weight less than 4 kg. Then $X \sim \text{Bin}\left(50, \frac{1}{3}\right)$.
2 Determine $E(X) = np$.	$E(X) = 50\left(\frac{1}{3}\right) = \frac{50}{3} = 16\frac{2}{3}$
3 Interpret your answer as a whole number of bags.	≈ 17 bags
d **1** Write the appropriate probability statement.	$P(X \geq 20) = P(20 \leq X \leq 50)$
2 Use CAS to evaluate the probability using **binomialCDf** for ClassPad and **binom Cdf** for TI-Nspire.	

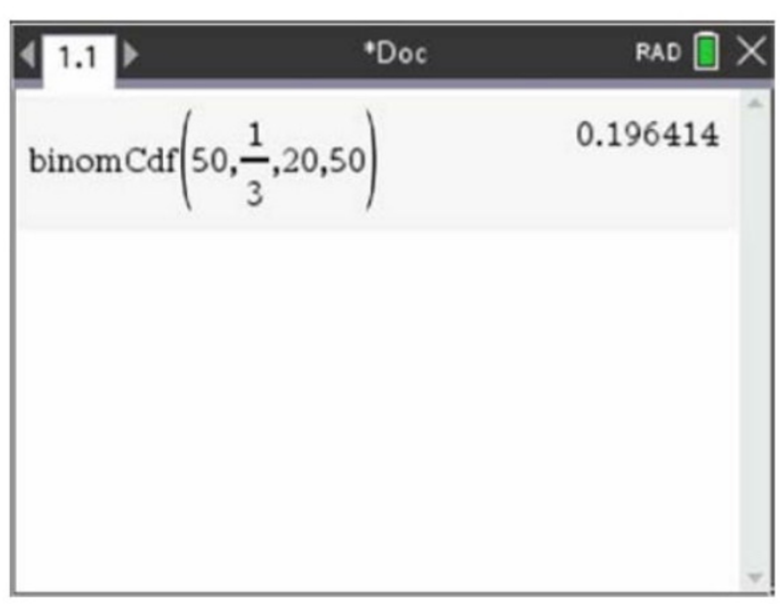
ClassPad





TI-Nspire





Steps	Working
3 Round to four decimal places.	$P(20 \leq X \leq 50) = 0.1964$

EXERCISE 8.3 Uniform and triangular distributions

ANSWERS p. 405

Recap

1 A continuous random variable has a probability density function given by

$$f(t) = \begin{cases} 0.05e^{-0.05t} & \text{for } t > 0 \\ 0 & \text{for } t \le 0 \end{cases}$$

The mean value of T is

A 0.05 **B** 2 **C** 5 **D** 10 **E** 20

2 A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{\pi}{2}\sin(\pi x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The value of the 50th percentile is

A $\frac{1}{2}$ **B** 1 **C** 1.5 **D** $\frac{\pi}{2}$ **E** π

Mastery

3 WORKED EXAMPLE 13 A uniformly distributed continuous random variable X has the probability density function

$$f(x) = \begin{cases} \frac{1}{6} & 9 \le x \le 15 \\ 0 & \text{otherwise} \end{cases}$$

Determine

a $P(X = 10)$ **b** $P(X > 10)$ **c** $P(X \le 12 | X > 10)$.

4 WORKED EXAMPLE 14 A uniformly distributed continuous random variable T has a cumulative distribution function given by

$$F(t) = \begin{cases} 0 & t < 7 \\ \frac{t-17}{13} & 7 \le t \le b \\ 1 & t > b \end{cases}$$

a Determine the value of b.

b State the probability density function of T.

c Hence, or otherwise, calculate $P(T > 15 | T > 10)$.

5 WORKED EXAMPLE 15 A uniformly distributed continuous random variable X is defined such that $X \sim U[-2, 48]$.

Find the

a probability density function of X

b expected value of X

c variance of X

d standard deviation of X.

6 WORKED EXAMPLE 16 Consider the following probability density function for the continuous random variable X.

$$f(x) = \begin{cases} \frac{1}{4}x - \frac{1}{2} & 2 \le x \le 4 \\ \frac{3}{2} - \frac{1}{4}x & 4 < x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(x)$ labelling the coordinates of all critical features. Hence, describe the shape of the distribution.

b Determine

i $P(X < 3)$ **ii** $P(X < 3 \mid X < 4)$ **iii** $E(X)$ **iv** $Var(X)$.

You may use your calculator for any calculations.

8.3

7 WORKED EXAMPLE 17 Consider the following probability density function for the continuous random variable X.

$$f(x) = \begin{cases} 2x & 0 \le x < \frac{1}{2} \\ -\frac{2}{3}x + \frac{4}{3} & \frac{1}{2} \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that it is a valid probability density function for X.

b Determine the 70th percentile.

c Show the use of integration to determine $E(X)$.

8 WORKED EXAMPLE 18 The lengths of plastic pipes that are cut by a particular machine are a uniformly distributed random variable, L, with lengths ranging from 245 mm to 251 mm. The machine is calibrated to cut pipes at a length of 250 mm.

a Determine the mean and standard deviation of L.

b Determine the probability that a randomly selected pipe has a length greater than the calibrated length of 250 mm.

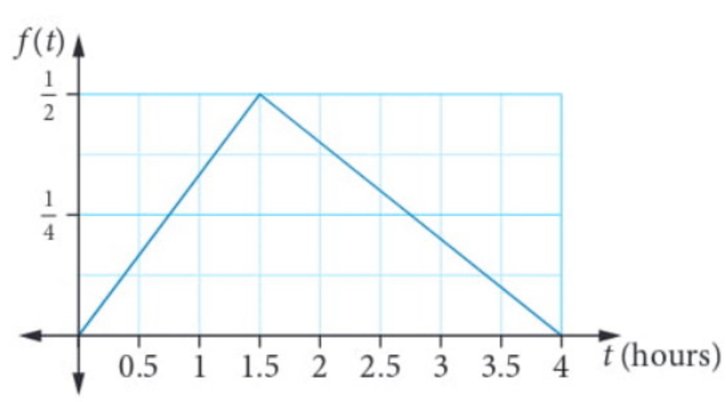
Suppose 60 different cut pipes are to be sampled.

c How many of the 60 pipes are expected to have a length greater than 250 mm?

d Determine the probability, correct to four decimal places, that at least 10 of the 60 pipes will have a length of more than 250 mm.

Calculator-free

9 © SCSA MM2019 Q3 MODIFIED (8 marks) Waiting times, T hours, for patients at a hospital emergency department can be up to four hours. The associated probability density function is shown below.



a Determine the equation of the probability density function, $f(t)$. (3 marks)

b What is the probability a patient will wait less than one hour? (2 marks)

c What is the probability a patient will wait between one hour and three hours? (3 marks)

10 © SCSA | MM2019 Q6 MODIFIED (11 marks) The error X in digitising a communication signal has a distribution with probability density function given by

$$f(x) = \begin{cases} 1 & -0.5 < x < 0.5 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $f(x)$ and, hence, describe the distribution. (3 marks)

b What is the probability that the error is at least 0.35? (1 mark)

c If the error is negative, what is the probability that it is less than −0.35? (2 marks)

d An engineer is more interested in the square of the error. What is the probability that the square of the error is less than 0.09? (2 marks)

e Calculate the variance of the error. (3 marks)

Calculator-assumed

11 (10 marks) The time Jennifer spends on her homework each day varies, but it is known that she does some homework every day. The continuous random variable T, which models the time, t, in minutes, that Jennifer spends each day on her homework, has a probability density function f, where

$$f(t) = \begin{cases} \frac{1}{625}(t-20) & 20 \le t < 45 \\ \frac{1}{625}(70-t) & 45 \le t \le 70 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(t)$, labelling the coordinates of all critical features. (3 marks)

b Hence, describe the shape of the distribution of T. (1 mark)

c Find

i $P(25 \le T \le 55)$ (2 marks)

ii $P(T \le 25 \mid T \le 55)$ (2 marks)

iii k such that $P(T \ge k) = 0.7$, correct to four decimal places. (2 marks)

12 (8 marks) In the MaxFun amusement park there is a small train called ChooChooCharlie which does a circuit of the park. The continuous random variable T, the time in minutes for a circuit to be completed, has a probability density function f with rule

$$f(t) = \begin{cases} \frac{1}{100}(t-10) & \text{if } 10 \le t < 20 \\ \frac{1}{100}(30-t) & \text{if } 20 \le t \le 30 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(t)$, labelling the coordinates of all critical features. (3 marks)

b Hence, describe the shape of the distribution of T. (1 mark)

c Find the exact probability that the time taken by ChooChooCharlie to complete a full circuit is

i less than 25 minutes (2 marks)

ii less than 15 minutes, given that it is less than 25 minutes. (2 marks)

8.3

13 © SCSA MM2016 Q16 (10 marks) An automated milk bottling machine fills bottles uniformly to between 247 mL and 255 mL. The label on the bottle states that it holds 250 mL.

a Determine the probability that a bottle selected randomly from the conveyor belt of this machine contains less than the labelled amount. (3 marks)

b Calculate the mean and standard deviation of the amount of milk in the bottles. (4 marks)

A worker selects bottles from the conveyor belt, one at a time.

c Determine the probability that in a selection of 15 bottles, five bottles containing less than the labelled amount have been selected. (3 marks)

14 (9 marks) Black Mountain coffee is sold in packets labelled 250 grams. Let the weight of a packet of Black Mountain coffee be represented by the continuous random variable W. The packing process produces packets whose weights form a uniform distribution over the interval $245 \leq w \leq 253$ grams.

a Determine the mean and standard deviation of W. (2 marks)

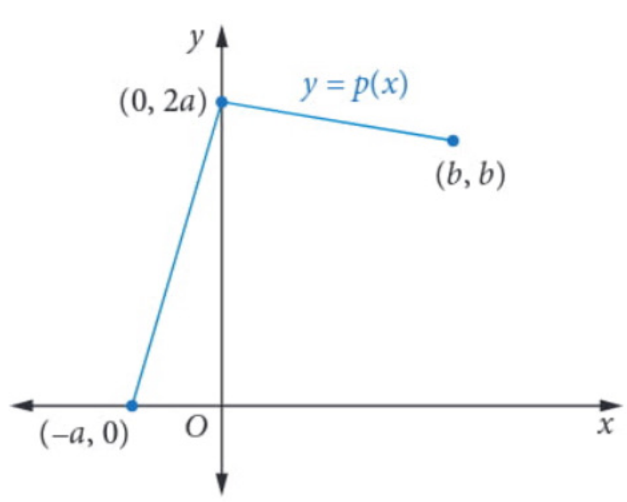
b Determine the probability that a randomly selected packet of Black Mountain coffee weighs more than the labelled weight of 250 g. (2 marks)

Suppose 100 different packets of Black Mountain coffee are to be sampled.

c How many of the 100 packets are expected to weigh more than 250 g? Answer to the nearest whole packet. (3 marks)

d Determine the probability, correct to four decimal places, that at least a quarter of the 100 packets weigh more than 250 g. (2 marks)

15 (8 marks) The distribution of a continuous random variable, X, is defined by the probability density function $p(x)$ over $-a \leq x \leq b$ where a and b are positive real constants. The graph of the function $y = p(x)$ is shown below.



It is known that the average value of y (i.e. the average *height* of the function) over the interval $-a \leq x \leq b$ is $\frac{3}{4}$.

a Write two equations in terms of a and b using the probability density function of X. (4 marks)

b Determine the exact values of a and b. (2 marks)

c Hence, determine $P(X > 0)$. (2 marks)

8.4 The normal distribution

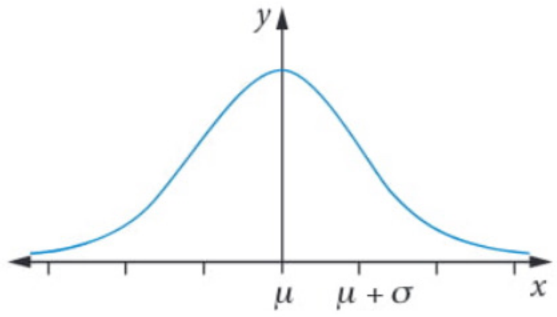
Video playlist
The normal distribution

Worksheet
The normal curve

Contexts suitable for the normal distribution

The **normal distribution** is one of the most frequently used probability distributions when modelling continuous random variables in both the natural and physical worlds, largely due to its critical features. A **normally distributed random variable** has

- a symmetrical distribution about its mean value, μ, meaning the mean, median and mode of a normal random variable are all equal
- a bell-shaped distribution, whereby the oblique points of inflection of the curve occur at values that are one standard deviation, σ, on either side of the mean.



Probability density function of a normally distributed random variable

For a normal random variable, X, with a mean μ and standard deviation σ, the probability density function is given by the equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The function $f(x)$ satisfies the two conditions of a continuous random variable:

- $f(x) > 0$ for all $-\infty < x < \infty$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

A normal random variable can be denoted by $X \sim N(\mu, \sigma^2)$, where the mean and variance are the parameters of the distribution.

The probability density function of a normal random variable can be used to show another critical feature of the normal distribution. This feature is known as the **68–95–99.7% rule**. These proportions indicate the probability that a value of x lies within one, two and three standard deviations from the mean, respectively.

The 68–95–99.7% rule

For $X \sim N(\mu, \sigma^2)$, then

- $P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$, i.e.

  $$\int_{\mu-\sigma}^{\mu+\sigma} f(x)\,dx \approx 0.68$$

- $P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$, i.e.

  $$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x)\,dx \approx 0.95$$

- $P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.997$, i.e.

  $$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x)\,dx \approx 0.997$$

This rule is useful to carry out approximate calculations without a calculator.

Exam hack

Sketch a bell-curve with a maximum of three standard deviations labelled either side of the mean to assist with probability calculations.

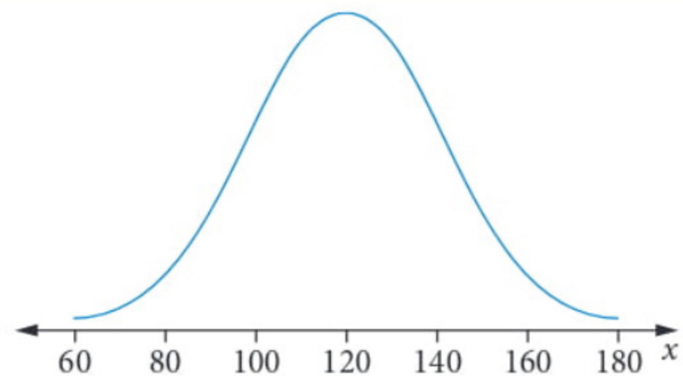
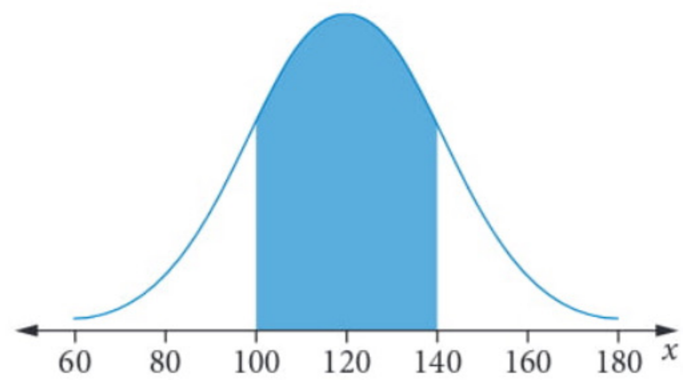
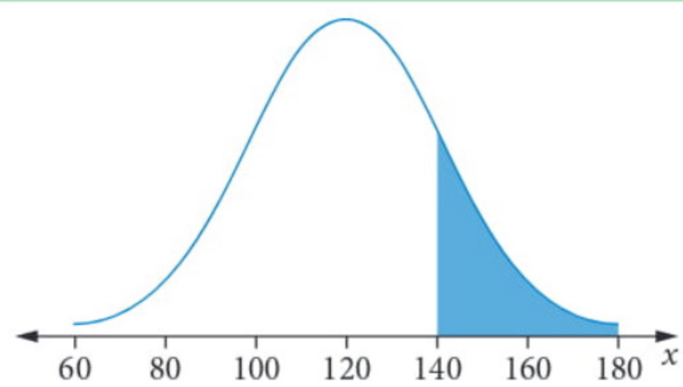
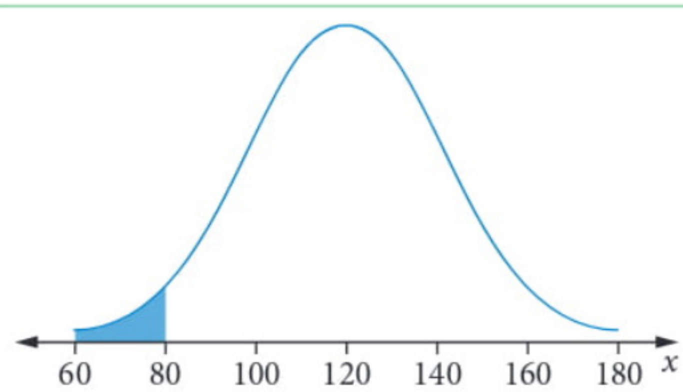
WORKED EXAMPLE 19 Using the 68–95–99.7% rule

A continuous random variable X is normally distributed with a mean of 120 and a standard deviation of 20. Use the 68–95–99.7% rule to approximate the following probabilities.

a $P(100 \le X \le 140)$

b $P(X > 140)$

c $P(X \le 80)$

Steps	Working
a **1** Sketch a **normal distribution curve** with a mean of 120 and standard deviation of 20.	
2 Shade the region corresponding to the bounds of 100 and 140, which are one standard deviation either side of the mean.	
3 Approximate the probability using the 68–95–99.7% rule.	$P(100 \le X \le 140) \approx 0.68$
b **1** Shade the region corresponding to the lower bound of 140, which is one standard deviation above the mean.	
2 Approximate the probability using the 68–95–99.7% rule.	$P(X \ge 120) = 0.5$ $P(120 \le X \le 140) \approx \frac{0.68}{2} \approx 0.34$ $P(X > 140) \approx 0.5 - 0.34 \approx 0.16$
c **1** Shade the region corresponding to the upper bound of 80, which is two standard deviations below the mean.	
2 Approximate the probability using the 68–95–99.7% rule.	$P(X \le 120) = 0.5$ $P(80 \le X \le 120) \approx \frac{0.95}{2} \approx 0.475$ $P(X < 80) \approx 0.5 - 0.475 \approx 0.025$

Together, the properties of

1 being a continuous variable, with
2 a symmetrical shape about the mean, such that
3 the proportions of scores either side of the mean follow the 68–95–99.7% rule

form the three conditions that can be used to determine whether a practical context can be suitably modelled by a normal distribution.

WORKED EXAMPLE 20 **Justifying contexts suitable for the normal distribution**

For each of the following situations, give one reason why it would **not** be appropriate to model the distribution of the continuous random variable with a normal distribution.

a The time taken (in minutes) for 100 students to complete an online road safety quiz has a mean of 25 minutes and standard deviation of 8 minutes. Seventy-five students completed the quiz within 17 to 33 minutes.

b A gardener has a nursery containing 2000 basil plants with a mean height of 14 cm and standard deviation of 4 cm. Approximately 1500 of the basil plants have a height, H, smaller than 14 cm.

c Results of a survey show that the number of electronic devices, D, owned by Australian teenagers is symmetrically distributed, with a mean number of devices of 2.

Steps	Working
Use the three properties of 1 continuous random variable 2 symmetrical shape 3 68–95–99.7% rule to find a reason why the context cannot be suitably modelled by a normal distribution.	**a** 1 Is time a continuous random variable? Yes 2 Is the distribution symmetrical? Insufficient information 3 Does the distribution satisfy the 68–95–99.7% rule? No $\mu - \sigma = 17$, $\mu + \sigma = 33$ $P(17 \leq T \leq 33) \approx 0.75 \neq 0.68$ Therefore, not appropriate as too many times are within one standard deviation from the mean.
	b 1 Is height a continuous random variable? Yes 2 Is the distribution symmetrical? No $\mu = 14$ $P(H < 14) = \dfrac{1500}{2000} = 0.75 \neq 0.5$ Therefore, not appropriate as the distribution is not symmetrical, i.e. it is skewed to the right (positively skewed), as there are more plants with a height less than the mean.
	c 1 Is number of devices a continuous random variable? No Therefore, not appropriate as the random variable is discrete.

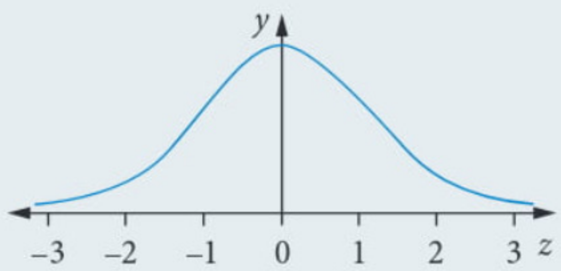
The standard normal distribution

Often a useful way to think about a normally distributed random variable, X, is to consider how many standard deviations a particular score x lies above or below the mean of the distribution $E(X) = \mu$. When normal distributions are thought about in this way, it does not actually matter what the values of the mean μ and standard deviation σ are, as we can set the mean as a default value of 0 and the standard deviation as 1 to obtain what is known as the **standard normal distribution**.

8.4

The standard normal distribution and z-scores

Let a normal random variable have a mean of 0 and a standard deviation of 1. This normal random variable is the **standard normal random variable** denoted by the distribution $Z \sim N(0, 1)$ and is used to describe the number of standard deviations a score is from the mean. This value is called a **z-score**.



The probability density function for a standard normal random variable is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Worksheets
The standard normal curve

Areas under the normal curve

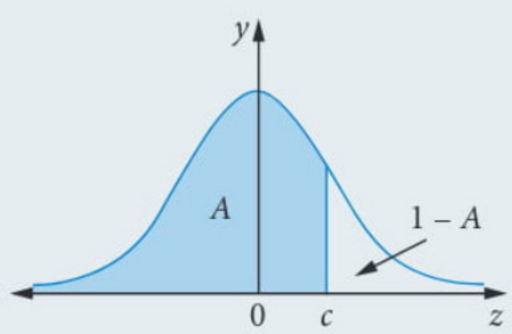
z-scores

From the 68–95–99.7% rule, we can then deduce that for $Z \sim N(0, 1)$:

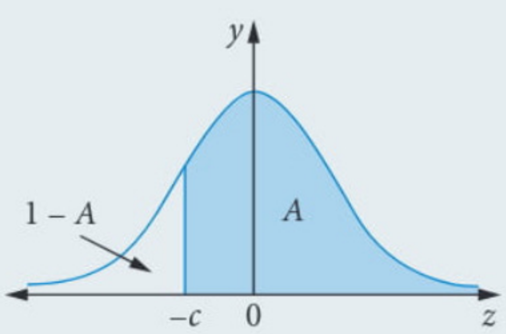
- $P(-1 \le Z \le 1) \approx 0.68$. That is, $\int_{-1}^{1} f(z)\,dz \approx 0.68$.
- $P(-2 \le Z \le 2) \approx 0.95$. That is, $\int_{-2}^{2} f(z)\,dz \approx 0.95$.
- $P(-3 \le Z \le 3) \approx 0.997$. That is, $\int_{-3}^{3} f(z)\,dz \approx 0.997$.

The symmetrical properties of the standard normal distribution

Let $P(Z \le c) = A$, then $P(Z > c) = 1 - A$, as the total area under the curve is 1.



By symmetry, $P(Z \ge -c) = A$ and so, $P(Z < -c) = 1 - A$.



WORKED EXAMPLE 21 **Solving problems involving $Z \sim N(0, 1)$**

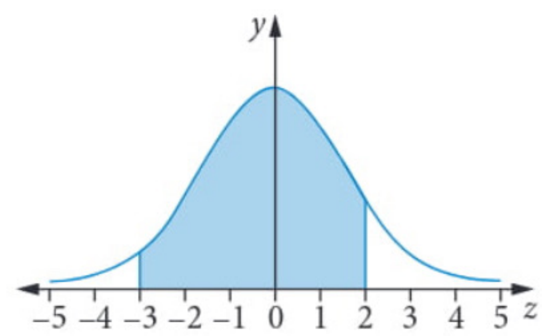
Let the continuous random variable Z be the standard normal random variable.

a Determine $P(-3 < Z < 2)$.

b If $P(Z \leq 1.5) = 0.933$ to three decimal places, determine $P(-1.5 \leq Z \leq 1.5)$ to three decimal places.

Steps | **Working**

a 1 Sketch the **standard normal distribution curve** and shade the region under the curve for $-3 < z < 2$.



2 Approximate the probability using the 68–95–99.7% rule.

$P(-3 < Z < 3) \approx 0.997$

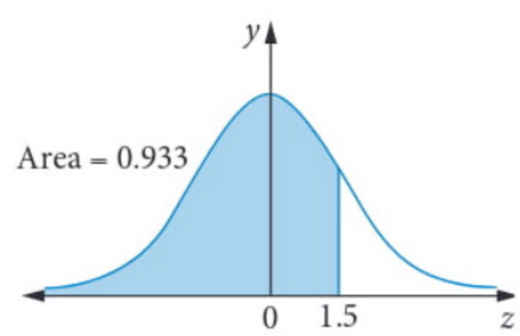
$P(-3 < Z < 0) \approx \dfrac{0.997}{2} \approx 0.4985$

$P(-2 < Z < 2) \approx 0.95$

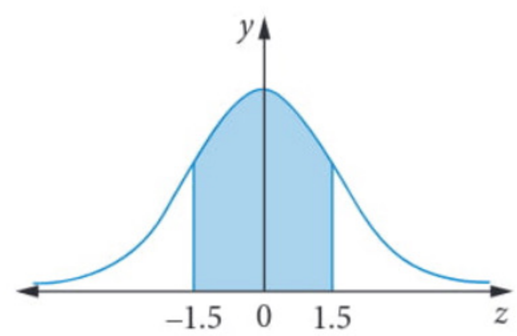
$P(0 < Z < 2) \approx \dfrac{0.95}{2} \approx 0.475$

$P(-3 < Z < 2) \approx 0.4985 + 0.475 \approx 0.9735$

b 1 Draw normal distribution curves that illustrate $P(Z \leq 1.5) = 0.933$ and $P(-1.5 \leq Z \leq 1.5)$.



$P(Z \leq 1.5) = 0.933$



$P(-1.5 \leq Z \leq 1.5)$

2 To find $P(-1.5 \leq Z \leq 1.5)$ (the region shaded under the second curve), first find $P(Z \geq 1.5)$ (the region unshaded under the first curve).

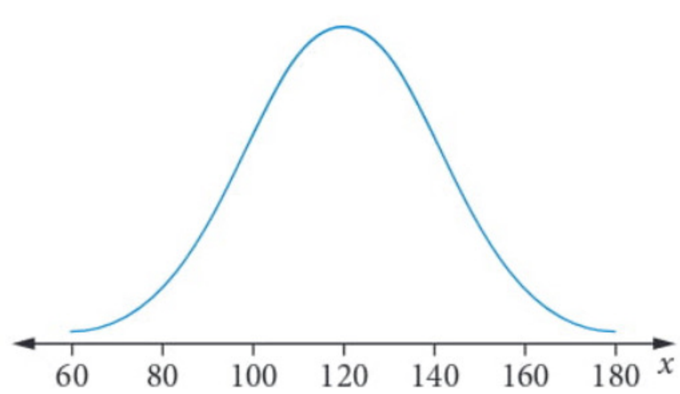
$P(Z \geq 1.5) = 1 - 0.933$
$= 0.067$

3 By symmetry, subtract double the value found above from 1.

$P(-1.5 \leq Z \leq 1.5) = 1 - 2 \times 0.067$
$= 0.866$

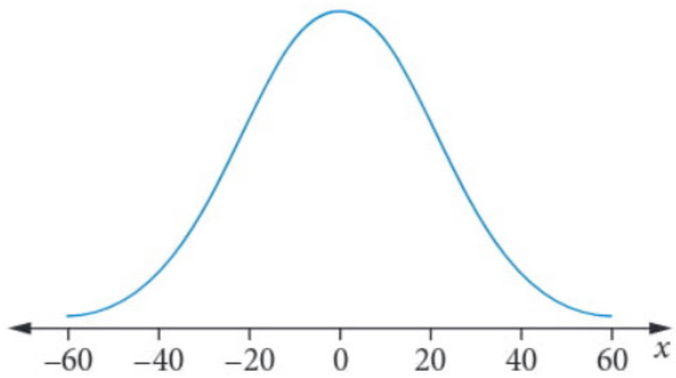
Whenever a normal random variable is not given in the standard normal form, that is, it is given as $X \sim N(\mu, \sigma^2)$ where $\mu \neq 0$ and $\sigma \neq 1$, it can always be scaled back to the standard normal distribution. That is, the value of μ can be made to be 0 and the value of σ can be made to be 1 such that all values of x are scaled back to the corresponding z-score (which is called its **standard score**). This process is called the **standardisation** of a normal random variable.

To do so, let's use the example $X \sim N(120, 20^2)$.

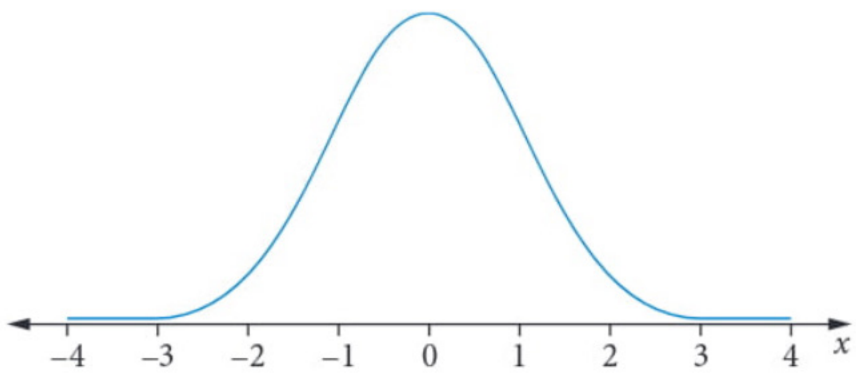


8.4

Imagine starting with the value of x that is the mean, $x = 120$. To make the mean 0, we would need to subtract the mean $\mu = 120$ from x, and to be consistent to the shape of the distribution, we would need to do it to every other value of x. That is, we apply a linear change of origin $X \to X - 120$.



Then we need the distance between the scores to reduce by a scale factor of the standard deviation $\sigma = 20$. That is, we need to apply a linear change of scale $X - 120 \to \frac{1}{20}(X - 120)$.



So, the standard normal random variable in this case would be defined as $Z = \frac{X - 120}{20}$.

Standardising a normal random variable

For a normal random variable $X \sim N(\mu, \sigma^2)$, the standard normal distribution $Z \sim N(0, 1)$ can be obtained using the linear transformation $Z = \frac{X - \mu}{\sigma}$.

To standardise a single value of x, that is, to obtain its z-score (standard score), use the formula

$$z = \frac{x - \mu}{\sigma}$$

where z is the standard score
x is the value of the score from the distribution $X \sim N(\mu, \sigma^2)$
μ is the mean of X
σ is the standard deviation of X.

It can then be shown that

$$\begin{aligned} E(Z) &= E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) \\ &= \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} \\ &= \frac{\mu}{\sigma} - \frac{\mu}{\sigma} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2}\mathrm{Var}(X) \\ &= \frac{1}{\sigma^2}(\sigma^2) \\ &= 1 \end{aligned}$$

When standardising a normal random variable, it is important to note that the corresponding areas under the curve **do not** change.

WORKED EXAMPLE 22 Standardising a normal random variable

A continuous random variable X is normally distributed with a mean of 80 and a standard deviation of 15. Let $Z \sim N(0, 1)$.

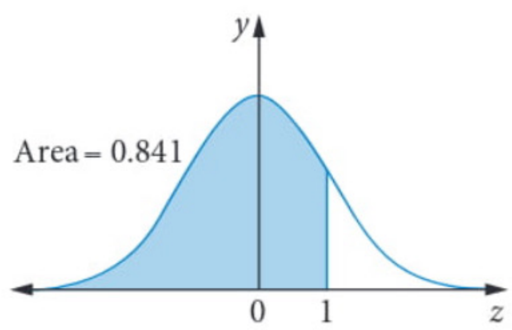
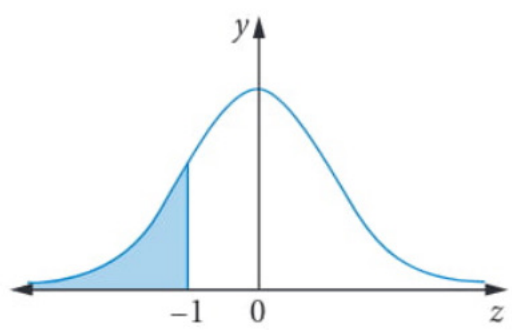
a Determine the standard score of the following values of x.

i $x = 65$

ii $x = 117.5$

iii $x = 77$

b If $P(Z \leq 1) = 0.841$, find $P(X < 65)$.

Steps	Working
a Use the z-score formula $z = \frac{x - \mu}{\sigma}$ for each of the values of x.	**i** $z = \frac{65 - 80}{15} = -\frac{15}{15} = -1$ **ii** $z = \frac{117.5 - 80}{15} = \frac{37.5}{15} = 2.5$ **iii** $z = \frac{77 - 80}{15} = \frac{-3}{15} = -0.2$
b **1** Draw a normal distribution curve that illustrates $P(Z \leq 1) = 0.841$.	 $P(Z \leq 1) = 0.841$
2 Use the standard score for $x = 65$ to draw a corresponding diagram of $P(X < 65)$ in terms of Z.	$x = 65 \Rightarrow z = -1$ (from part **a i**) 
3 Calculate the probability using symmetry.	$P(Z > -1) = P(Z \leq 1) = 0.841$ $P(X < 65) = P(Z < -1) = 1 - P(Z > -1) = 1 - 0.841 = 0.159$

Calculating probabilities and quantiles with the normal distribution

8.4

Although the 68–95–99.7% rule and z-scores are useful in helping us to calculate probabilities that are a 'nice' number of standard deviations from the mean, we will typically need the assistance of CAS to compute any other probabilities.

The ClassPad and the TI-Nspire use the **normCDf** and **Normal Cdf** respectively function to carry out probabilities of the form $P(a \le X \le b)$. When dealing with problems such as $P(X \le b)$ or $P(X \ge a)$, you will need to think about these in terms of $\pm\infty$ for your calculator input. That is:

- $P(X \le b) = P(-\infty < X \le b)$
- $P(X \ge a) = P(a \le X < \infty)$.

Worksheets
The standard normal curve
The normal distribution

USING CAS 4 Probabilities for a normally distributed random variable

A continuous random variable X is normally distributed with mean 20 and standard deviation 5.
Find the following probabilities, correct to four decimal places.

a $P(X \le 28)$

b $P(X > 19)$

c $P(15 < X < 22)$

ClassPad

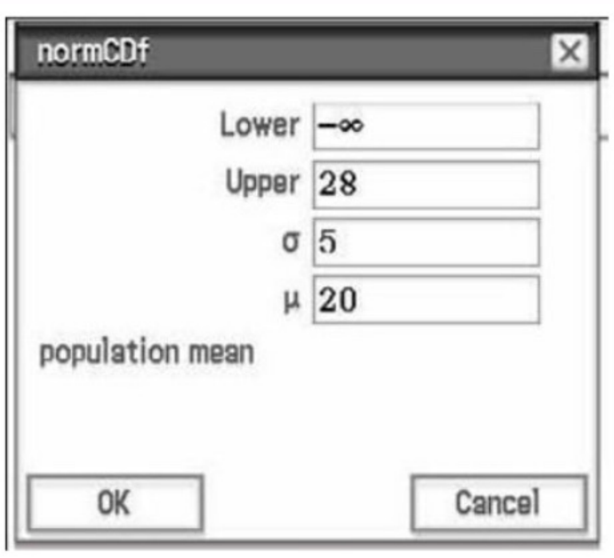
1 Tap **Interactive** > **Distribution/Inv. Dist** > **Continuous** > **normCDf**.

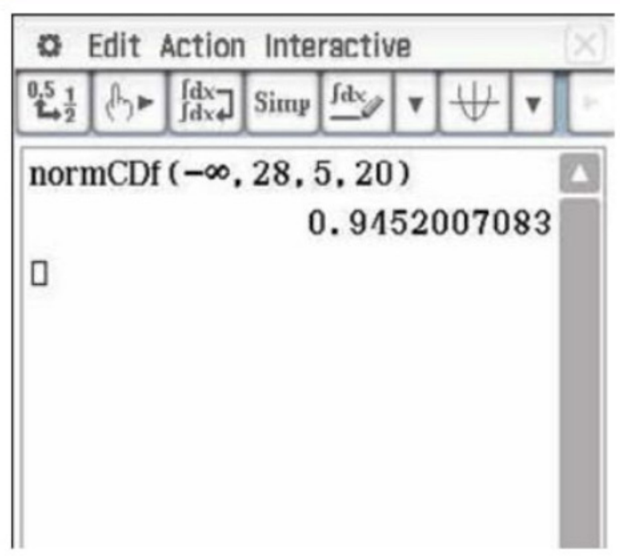
2 In the dialogue box, enter the corresponding lower, upper, σ and μ values as shown.

3 Tap **OK** and the probability will be displayed.

(Note: always use $-\infty$ or ∞ from the **Math2 keyboard** for the 'ends' of the normal distribution.)

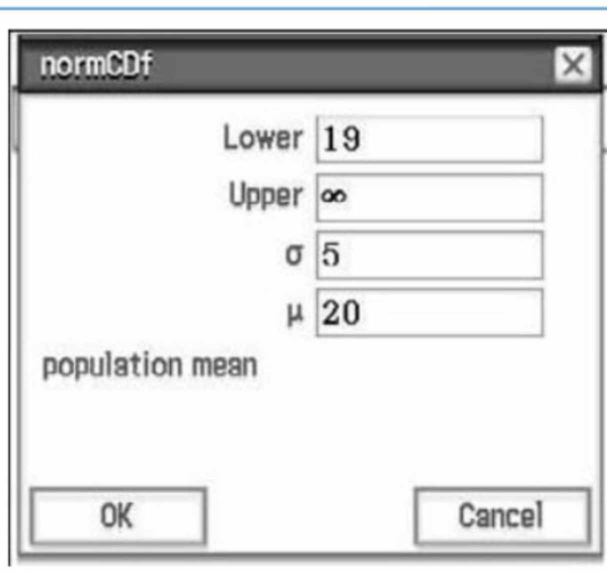
a

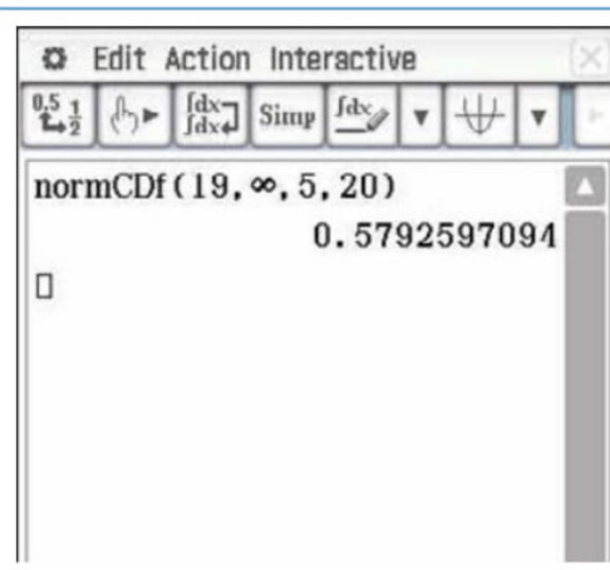




$P(X \le 28) = 0.9452$

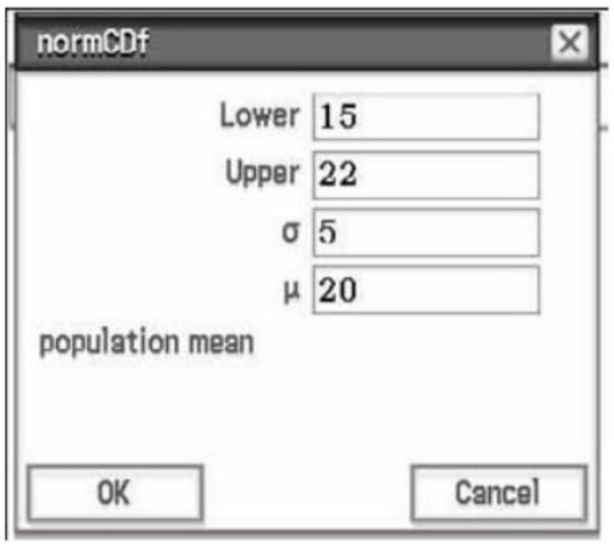
b

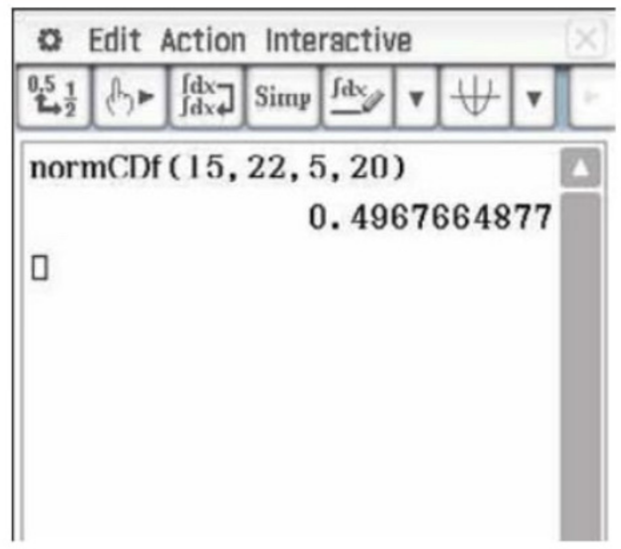




$P(X > 19) = 0.5793$

c





$P(15 < X < 22) = 0.4968$

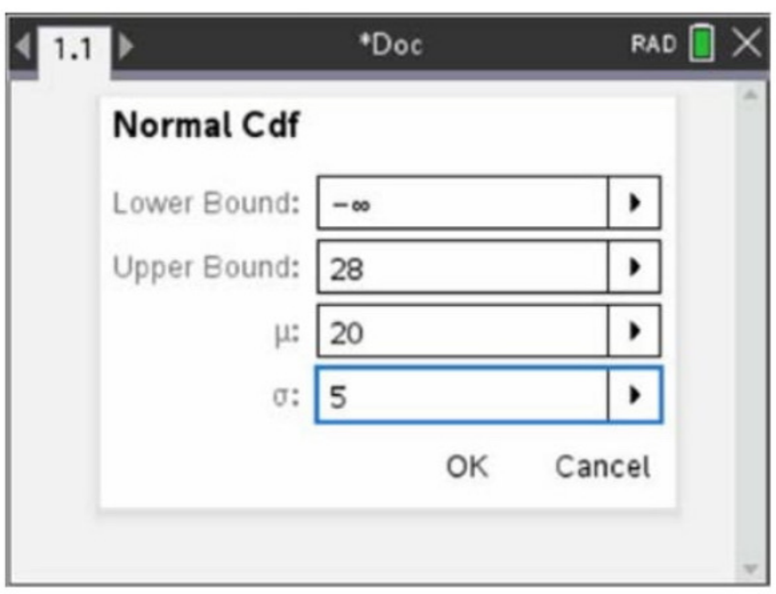
TI-Nspire

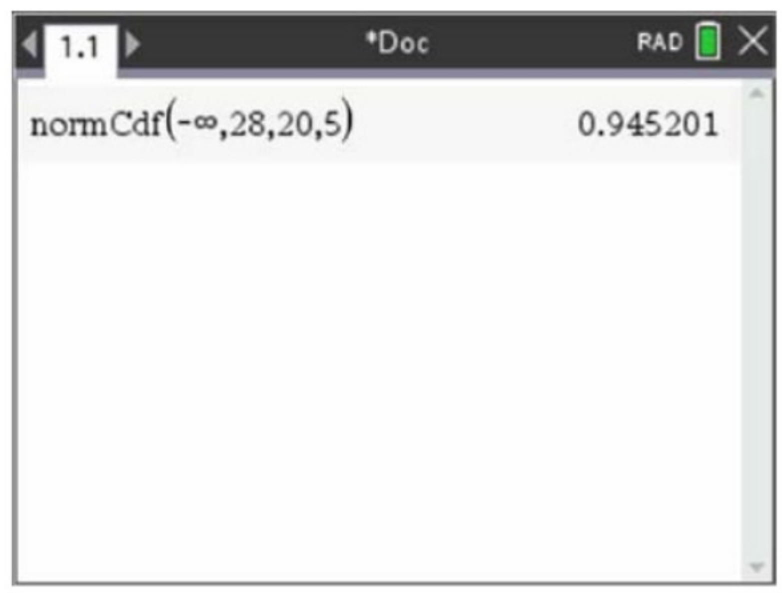
1 Press **menu** > **Probability** > **Distributions** > **Normal Cdf**.

2 In the dialogue box, enter the corresponding lower bound, upper bound, μ and σ values as shown.

3 The probability will be displayed.

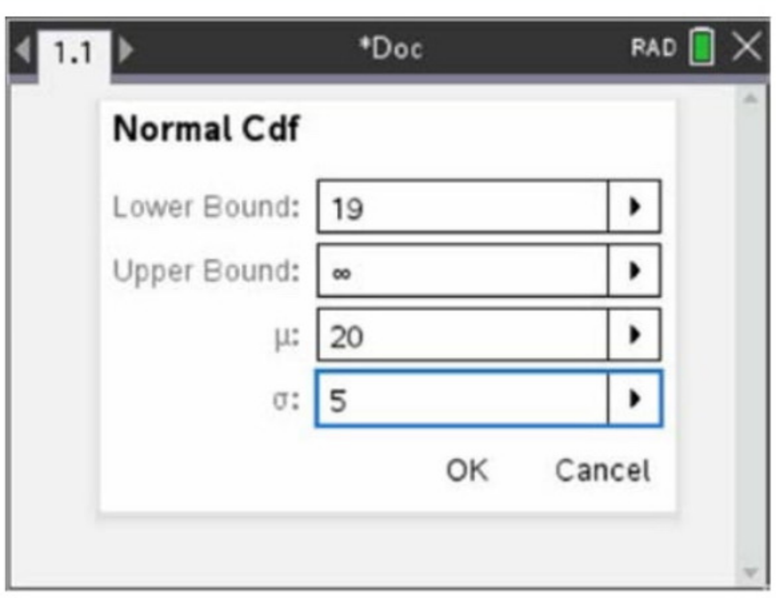
a

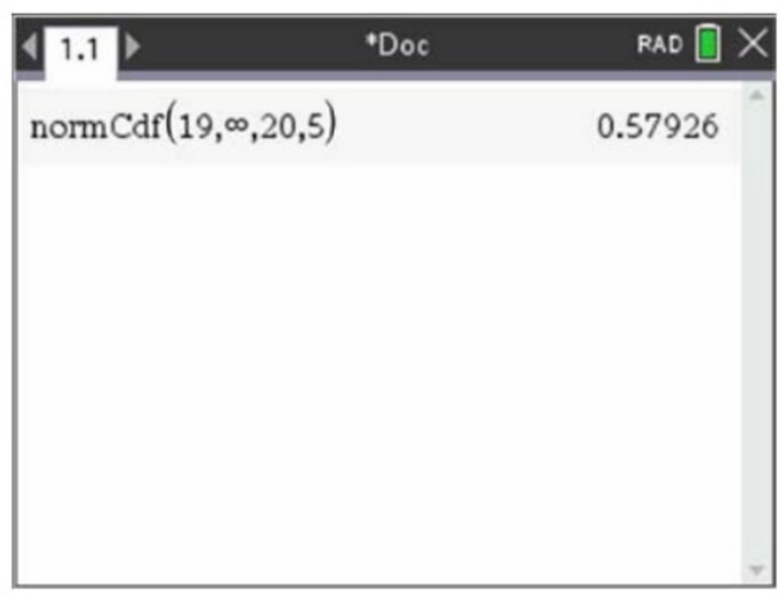




$P(X \leq 28) = 0.9452$

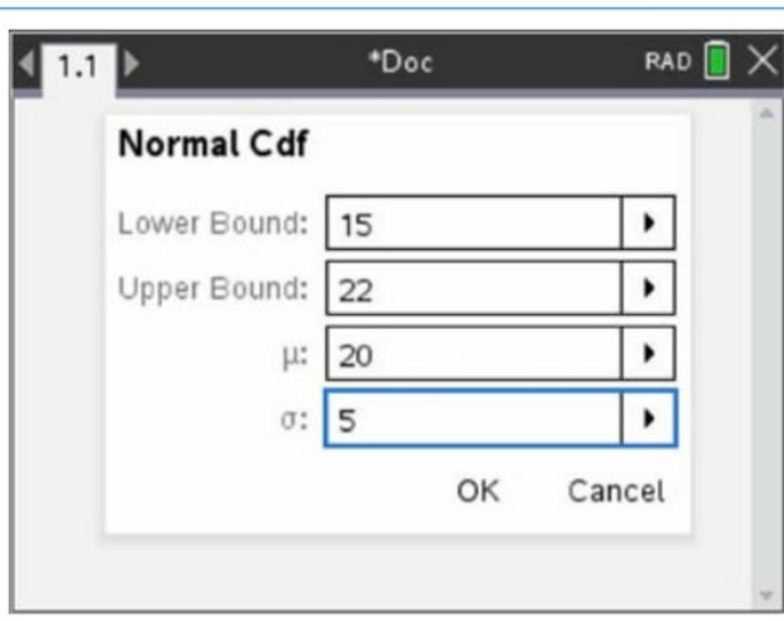
b

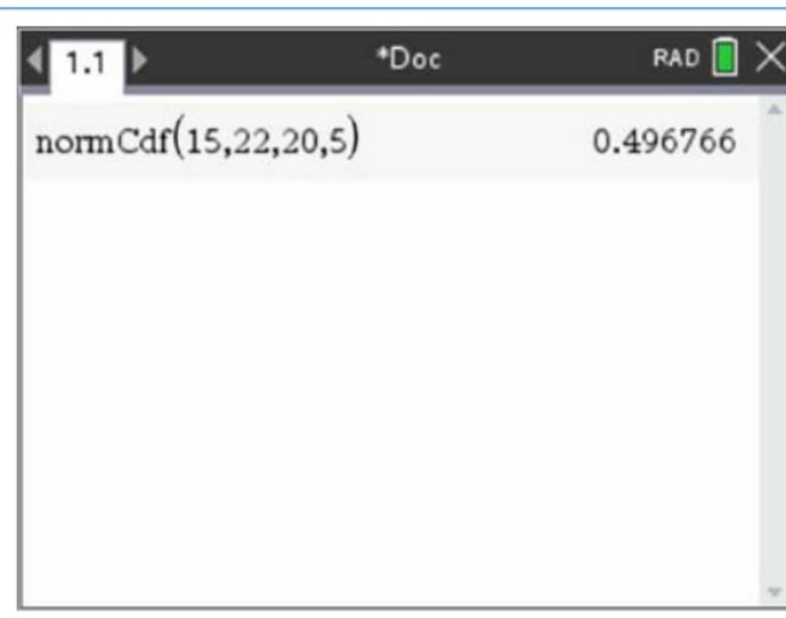




$P(X > 19) = 0.5793$

c





$P(15 < X < 22) = 0.4968$

9780170477536

As with general continuous random variables, we can solve problems of the form $\mathrm{P}(X \le k) = p$ with normal random variables. Previously we saw the language of the percentile to describe such a problem. For example, the 75th percentile is the value of k such that $\mathrm{P}(X \le k) = 0.75$. This value of k can also be referred to as the 0.75 **quantile**.

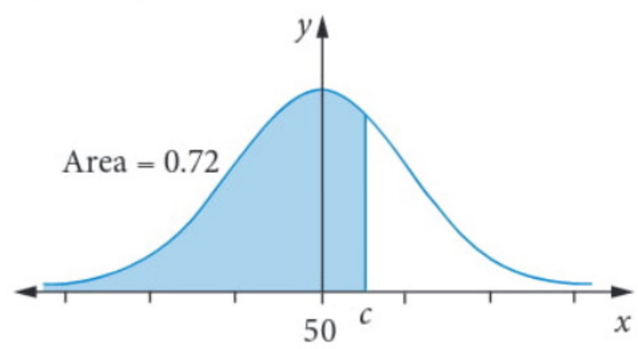
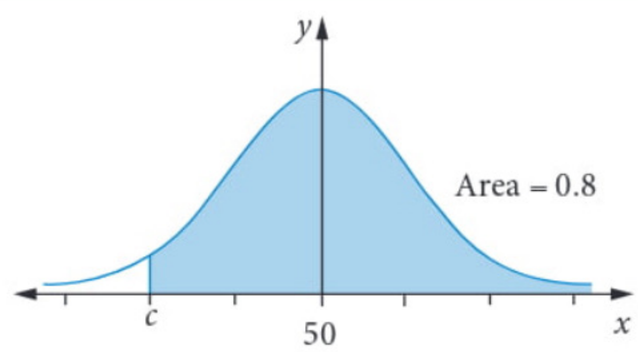
A quantile is simply the decimal form of a percentile, such that the p quantile is the score below which that $100p\%$ of the variable lies, where $0 < p < 1$.

Once again, we can use CAS and the **inverse normal distribution** function to solve these problems, but the ClassPad and TI-Nspire use slightly different conventions.

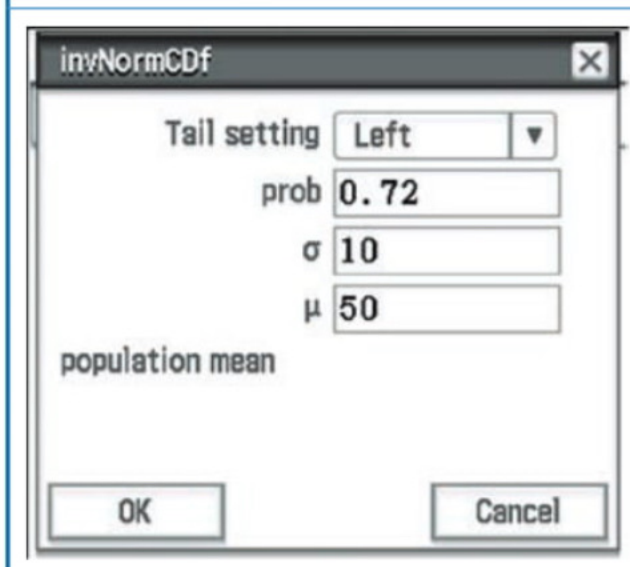
USING CAS 5 Finding quantiles using the inverse normal distribution

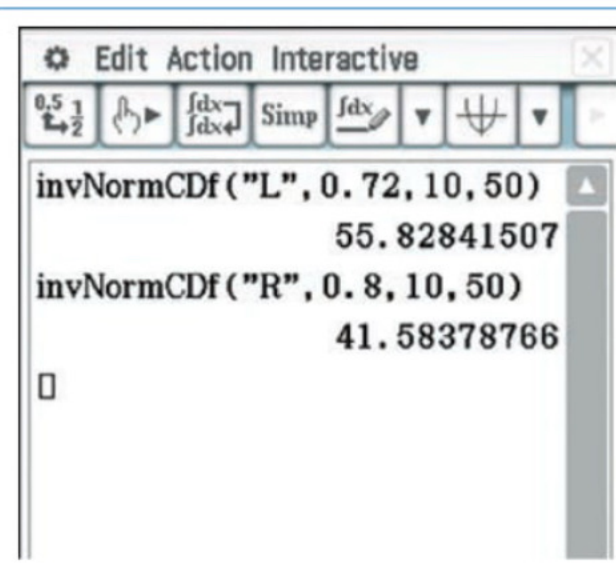
A continuous random variable X is normally distributed with a mean of 50 and a standard deviation of 10. Find the value of c, correct to two decimal places, if

a $\mathrm{P}(X \le c) = 0.72$ **b** $\mathrm{P}(X \ge c) = 0.8$.

Steps	Working
a 1 Draw the normal distribution curve, label c on the x-axis, and shade the given area. $\mathrm{P}(X \le c) = 0.72$ indicates that 72% of the values are less than c. **2** Use CAS to solve for the 0.72 quantile.	$\mu = 50$, $\sigma = 10$ 
b 1 Draw the normal distribution curve, label c on the x-axis, and shade the given area. $\mathrm{P}(X \ge c) = 0.8$ indicates that 80% of the values are greater than c. **2** Use CAS to solve for the 0.20 quantile.	

ClassPad





1 Tap **Interactive** > **Distribution/Inv.Dist** > **Inverse** > **invNormCDf**.

2 In the dialogue box, use the default **Tail setting** as **Left** for < or ≤.

3 Enter the values as shown above.

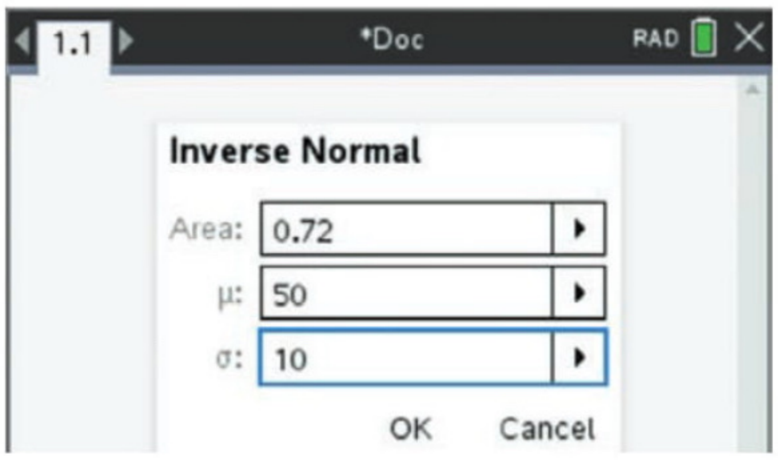
4 The answer for part **a** will be displayed.

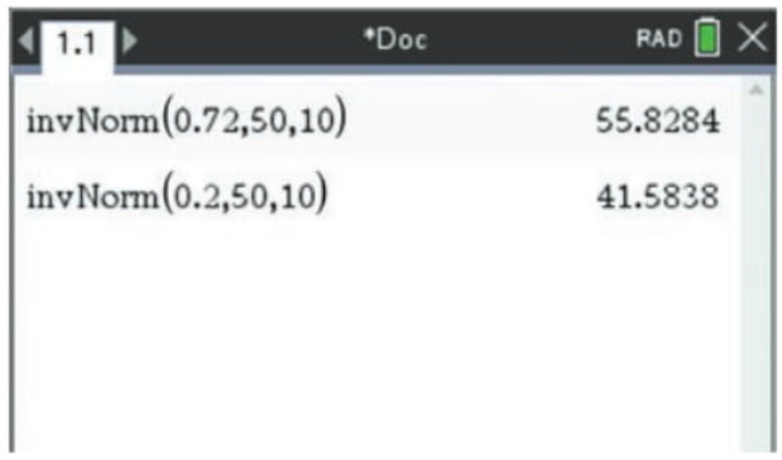
5 Repeat for part **b** but change the **Tail** setting to **Right**.

a $c \approx 55.83$

b $c \approx 41.58$

TI-Nspire





1 Press **menu** > **Probability** > **Distributions** > **Inverse Normal**.

2 In the dialogue box, enter the values as shown above.

3 The answer for part **a** will be displayed.

4 Repeat for part **b** using **area = 1 − 0.8 = 0.2**.

a $c \approx 55.83$

b $c \approx 41.58$

Worksheets
Applying the normal distribution
Normal distribution – Worded problems 1
Normal distribution – Worded problems 2

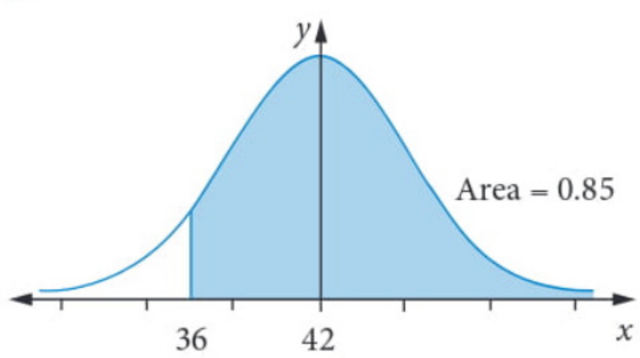
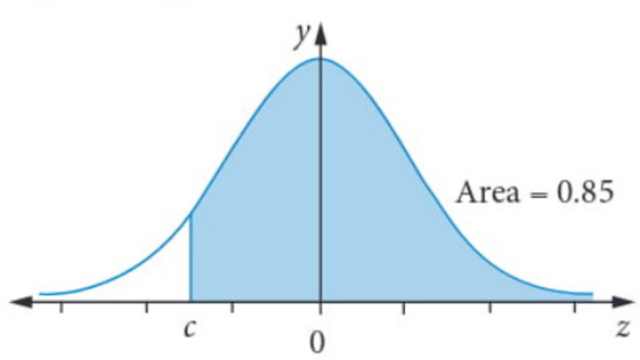
Problems involving the normal and binomial distributions

In some problems, we may not have all the information about the parameters of a normal random variable; that is, either μ or σ^2 or both may be unknown. In those cases, we must have sufficient information about probabilities or corresponding z-scores in order to determine the unknown parameters.

Typically for these problems, we should look to connect the values of μ and σ with x and its corresponding z-score using the formula $z = \dfrac{x - \mu}{\sigma}$.

WORKED EXAMPLE 23 Finding a parameter of a normal distribution

A continuous random variable X has the distribution $X \sim N(42, \sigma^2)$. If $P(X \geq 36) = 0.85$, find the standard deviation of X, correct to two decimal places.

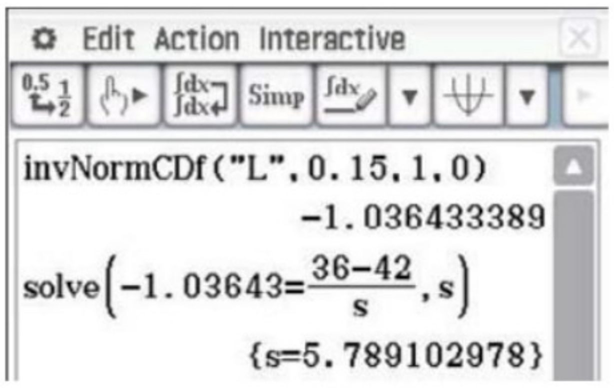
Steps	Working
1 Draw the normal distribution curve for X and the corresponding normal distribution curve for Z.	$\mu = 42$  $P(X \geq 36) = 0.85$  $P(Z \geq c) = 0.85$ $P(Z < c) = 1 - 0.85 = 0.15$
2 Use the CAS inverse normal distribution to find c (see Step 3 on the following page).	$c = -1.036$

3 Substitute into $z = \frac{x-\mu}{\sigma}$ and solve for σ algebraically or using CAS.

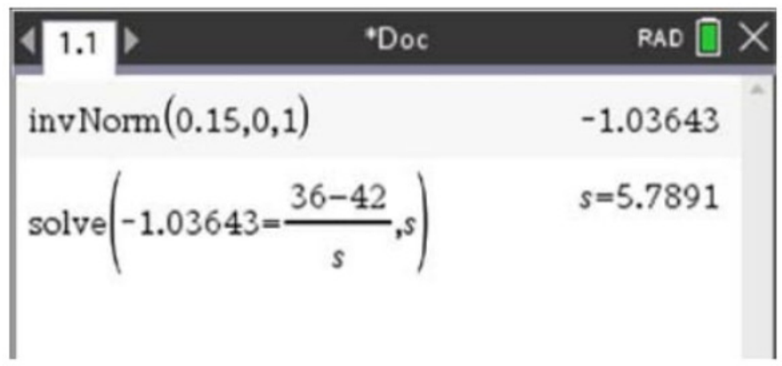
$\mu = 42, x = 36, z = -1.036$

$$z = \frac{x-\mu}{\sigma}$$
$$-1.036 = \frac{36-42}{\sigma}$$
$$-1.036\sigma = -6$$
$$\sigma = 5.79$$

ClassPad



TI-Nspire



Be prepared to solve problems involving normal random variables in contextual situations, incorporating all the knowledge and skills you have learnt in this exercise. Like with the uniform and triangular distributions, be prepared to revisit the use of the binomial distribution in situations whereby you have n trials and a probability of success p being obtained from a normal distribution.

WORKED EXAMPLE 24 Applying the normal distribution in context

Year 10 students complete a fitness endurance task. The times taken to complete the task are normally distributed with a mean of 15 minutes and a standard deviation of 2 minutes.

a Find, correct to four decimal places, the proportion of students who complete the task in less than 14 minutes.

b Students who complete the task in a time between 13.5 minutes and 17 minutes are classified as having average fitness levels. Find, correct to four decimal places, the probability of a student selected at random being classified as having average fitness.

c Find, correct to four decimal places, the probability of a student with average fitness completing the task in less than 14 minutes.

Suppose a sample of Year 10 students was to be taken, and 30 of these students classified as having average fitness.

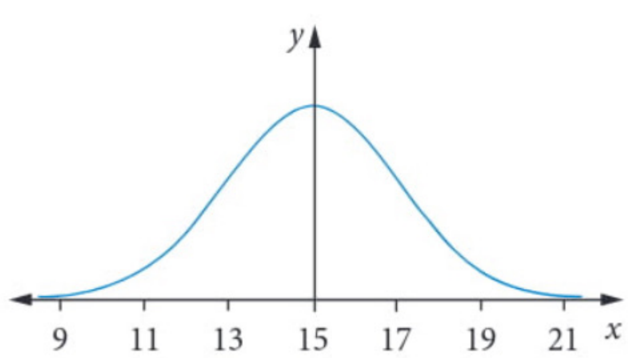
d Determine the probability, correct to four decimal places, that exactly 3 of the 30 students of average fitness completed the task in less than 14 minutes.

Steps | **Working**

a 1 Define an appropriate random variable for the context.

Let X represent the time taken to complete the fitness task in minutes, such that $X \sim N(15, 4)$.

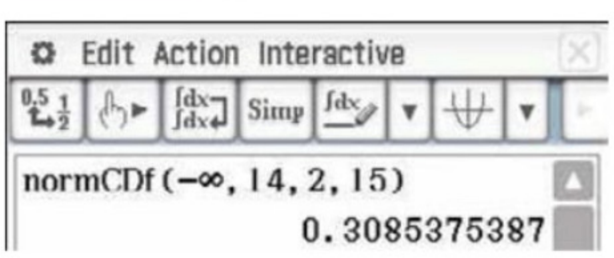
2 Draw the normal distribution curve and show the three standard deviations above and below the mean on the horizontal axis scale.



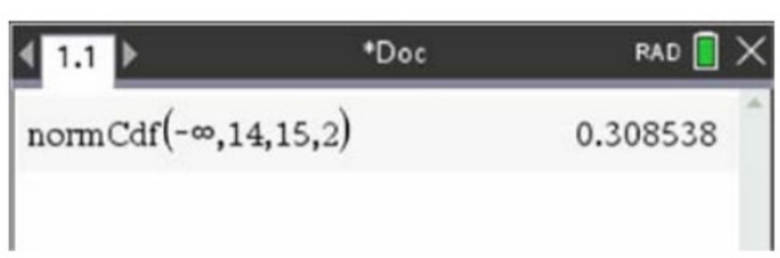
3 Calculate $P(X < 14)$ using CAS.

$P(X < 14) = 0.3085$

ClassPad



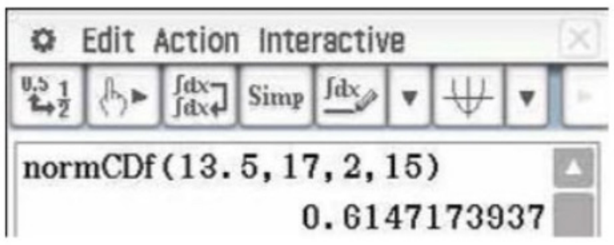
TI-Nspire



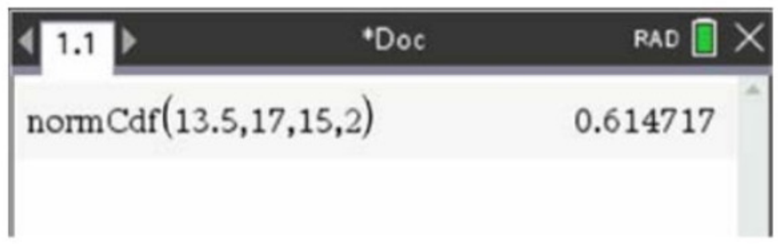
b Calculate $P(13.5 < X < 17)$ using CAS.

$P(13.5 < X < 17) = 0.6147$

ClassPad



TI-Nspire



c **1** Recognise the conditional language 'student with average fitness completing the task in less than 14 minutes' as 'a student completes a task in under 14 minutes given they have average fitness'.

2 Write and use the conditional probability formula $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$.

$P(X < 14 \mid 13.5 < X < 17)$

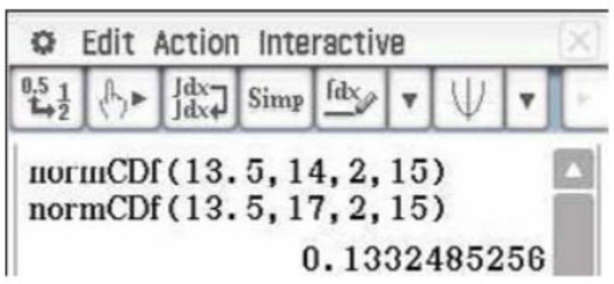
$= \dfrac{P(13.5 < X < 14)}{P(13.5 < X < 17)}$

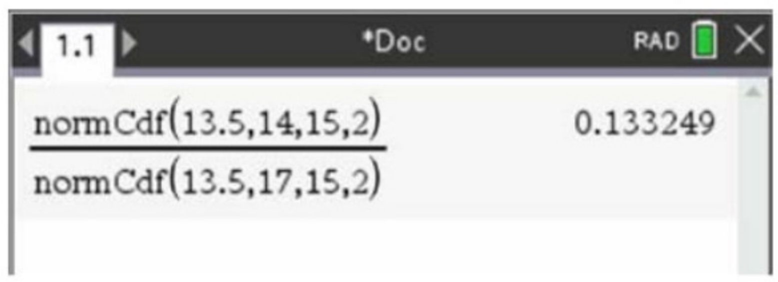
3 Calculate using CAS.

$= \dfrac{P(13.5 < X < 14)}{P(13.5 < X < 17)}$

$= \dfrac{0.081\,9102\ldots}{0.614\,7174\ldots}$

$= 0.1332$





d **1** Define a new binomial random variable and its distribution.

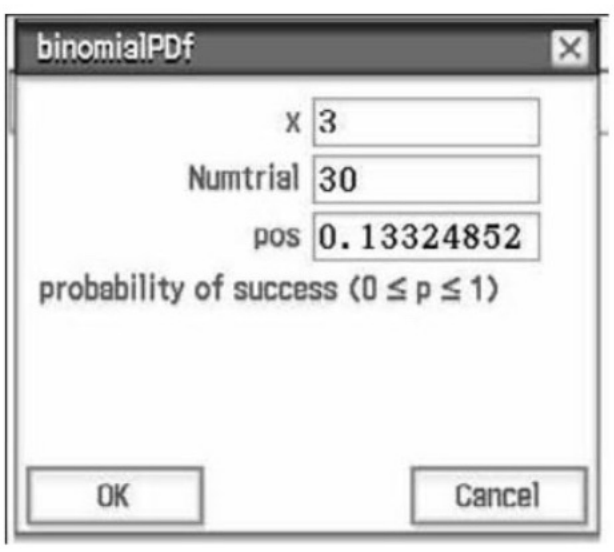
Let Y be the number of students out of 30 who completed the task in less than 14 minutes, given that they classified as having average fitness.

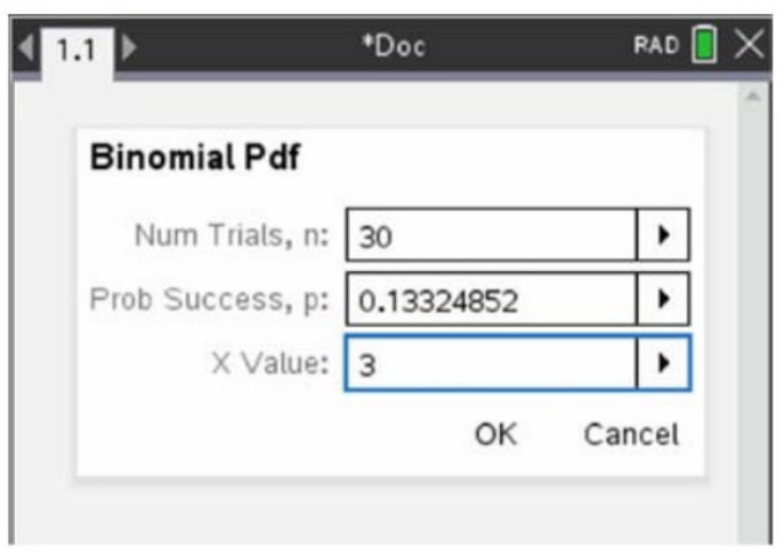
$Y \sim \text{Bin}(30, 0.1332)$

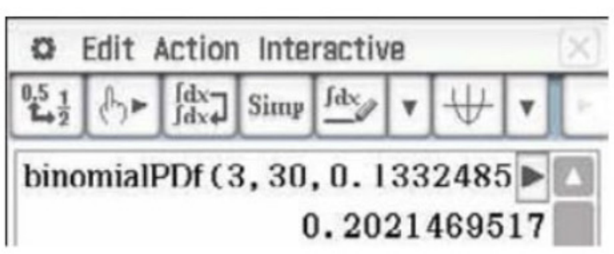
2 Establish the **binomialPDf** (ClassPad) or **Binomial Pdf** (TI-Nspire) calculation and use CAS to solve.

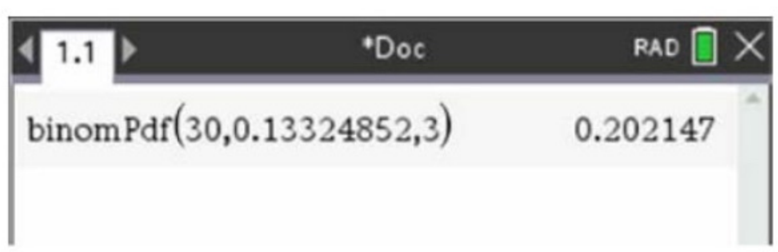
$P(Y = 3) = \binom{30}{3}(0.1332)^3(1 - 0.1332)^{27}$

$= 0.2021$









Exam hack

In questions involving probabilities that carry through multiple parts, be sure to use the full values from CAS.

WACE QUESTION ANALYSIS

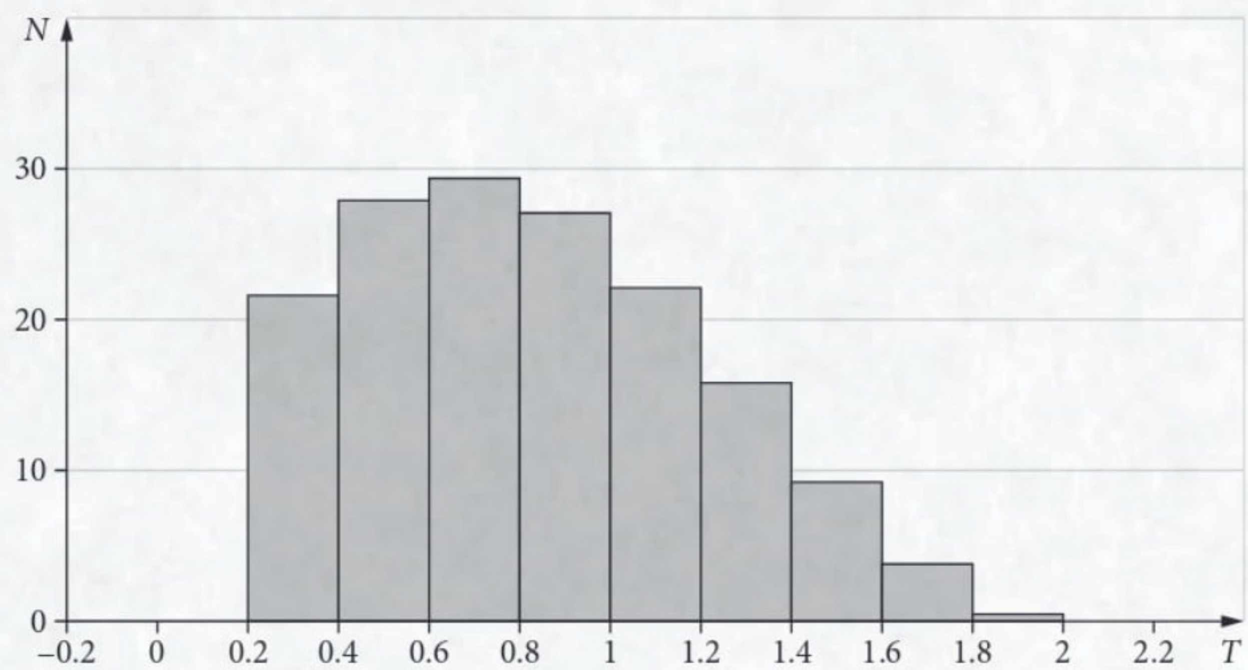
 MM2020 Q16 Calculator-assumed (7 marks)

A large refrigerator in a scientific laboratory is always required to maintain a temperature between 0°C and 1°C to preserve the integrity of biological samples stored inside. A scientist working in the laboratory suspects that the refrigerator is not maintaining the required temperature and decides to record the temperature every hour for seven days. Based on these measurements, the scientist concludes that the temperature, T, in the refrigerator is normally distributed with a mean of 0.8°C and a standard deviation of 0.4°C.

a Temperature in degrees Fahrenheit, T_f, is given by $T_f = \frac{9}{5}T + 32$. Determine the mean and standard deviation of the refrigerator temperature in degrees Fahrenheit. (2 marks)

b Determine the probability that the refrigerator temperature is above 1°C. Give your answer rounded to four decimal places. (1 mark)

The histogram of data gathered by the scientist is shown below. N denotes the number of observations in each temperature interval.



c Do you agree that the normal distribution was an appropriate model to use? Provide a reason to justify your response. (2 marks)

An alternative probability density function proposed to model the refrigerator temperature, in degrees Celsius, is given by:

$$p(t) = \frac{3}{4}t^3 - 3t^2 + 3t, \ 0 \le t \le 2$$

d Determine the probability that the refrigerator temperature is above 1°C using the new model. (2 marks)

Reading the question

- Identify and highlight the type of distribution and its corresponding parameters.
- Highlight any rounding instructions for probability questions.
- Highlight any key command words, e.g. *justify*.

Thinking about the question

- Consider the properties of change of scale and origin for $E(aX + b)$ and $SD(aX + b)$.
- Make a mental list of the conditions that make a normal distribution a suitable model for a continuous random variable.

Video
WACE question analysis: Continuous random variables and the normal distribution

Worked solution (✓ = 1 mark)

a The mean of T_f is

$$\mu_{T_f} = \frac{9}{5}\mu_T + 32$$
$$= \frac{9}{5}\left(\frac{4}{5}\right) + 32$$
$$= \frac{836}{25} = 33\frac{11}{25} = 33.44$$

The standard deviation of T_f is

$$\sigma_{T_f} = \frac{9}{5}\sigma_T$$
$$= \frac{9}{5}\left(\frac{2}{5}\right)$$
$$= \frac{18}{25} = 0.72$$

correctly determines mean using a change of scale and origin ✓

correctly determines standard deviation using a change of scale ✓

b $T \sim N(0.8, 0.4^2)$

$P(T > 1) = 0.3085$

determines the correct probability using CAS ✓

c No. The distribution appears to be skewed to the right (non-symmetric).

states that it is not an appropriate model ✓

justifies conclusion based on the lack of symmetry in the histogram ✓

d
$$P(T \geq 1) = \int_1^2 p(t)\,dt$$
$$= \int_1^2 \left(\frac{3}{4}t^3 - 3t^2 + 3t\right)dt$$
$$= \left[\frac{3}{16}t^4 - t^3 + \frac{3}{2}t^2\right]_1^2$$
$$= 1 - \frac{11}{16}$$
$$= \frac{5}{16} \quad \{0.3125\}$$

establishes the correct integral to determine the probability ✓

evaluates the integral to obtain the correct probability ✓

Exam hack

Be sure to pay attention to rounding instructions within questions, as they form part of the marking behaviours! In all answers requiring a written response, be sure to communicate clearly with an appropriate use of mathematical terminology; for example, correct use of the terms skewness and/or symmetry.

EXERCISE 8.4 The normal distribution

ANSWERS p. 407

8.4

Recap

Questions 1 and 2 relate to the context below.

The lifetime T, in hours, of a particular type of light globe can be modelled by a symmetrical triangular distribution over the interval $50 \le t \le 150$.

1 The mean lifetime of the light globes in hours is

A 1 **B** 100 **C** 500 **D** 800 **E** 1000

2 The maximum value of the probability density function $f(t)$ is

A 0.01 **B** 0.02 **C** 0.2 **D** 1 **E** 2

Mastery

3 WORKED EXAMPLE 19 A continuous random variable X is normally distributed with a mean of 62 and a standard deviation of 8. Use the 68–95–99.7% rule to approximate the following probabilities.

a $P(46 \le X \le 78)$ **b** $P(X > 70)$ **c** $P(X \le 38)$

4 WORKED EXAMPLE 20 For each of the following situations, give one reason why it would **not** be appropriate to model the distribution of the continuous random variable with a normal distribution.

a The mass, M, of a collection of different vehicles has a mean of 1500 kg and a standard deviation of 700 kg.

b The number of followers, F, that students in a class have on their Snaptagram accounts has a mean of 7 and a standard deviation of 1.9.

5 WORKED EXAMPLE 21 Let the continuous random variable Z be the standard normal random variable.

a Determine $P(-1 < Z < 2)$.

b If $P(Z \le a) = 0.82$ and $P(Z \le b) = 0.18$ to two decimal places, determine $P(b \le Z \le a)$.

6 WORKED EXAMPLE 22 A continuous random variable X is normally distributed with a mean of 35 and a standard deviation of 7. Let $Z \sim N(0, 1)$.

a Determine the standard score of the following values of x.

i $x = 14$ **ii** $x = 49$ **iii** $x = 59.5$

b If $P(Z \ge -2) = 0.975$, determine $P(X \ge 49)$.

7 Using CAS 4 A continuous random variable X is normally distributed with mean 150 and standard deviation 15. Find the following probabilities, correct to four decimal places.

a $P(X > 188)$ **b** $P(X \le 140)$ **c** $P(132 < X < 159)$

8 Using CAS 5 A continuous random variable X is normally distributed with a mean of 2400 and a standard deviation of 400. Find the value of c, correct to two decimal places, if

a $P(X \le c) = 0.28$ **b** $P(X \ge c) = 0.65$.

9 WORKED EXAMPLE 23 A continuous random variable X has the distribution $X \sim N(\mu, 50^2)$. If $P(X \le 170) = 0.2743$, find the mean of X correct to the nearest integer.

10 WORKED EXAMPLE 24 The Clucky Hen Egg Farm produces eggs whose weights are normally distributed with a mean of 78 g and a standard deviation of 6 g.

a Find the probability, correct to four decimal places, that a randomly selected egg weighs more than 69 g.

b Eggs that have a weight larger than 80 g are considered 'Jumbo' eggs. Find, correct to four decimal places, the probability of an egg selected at random being classified as a 'Jumbo' egg.

c Find, correct to four decimal places, the probability of an egg weighing more than 69 g being classified as 'Jumbo'.

Suppose a sample of eggs was taken, with 80 eggs found to weigh more than 69 g.

d Determine the probability, correct to four decimal places, that exactly half of these 80 eggs will be classified as 'Jumbo'.

Calculator-free

11 (3 marks) Let X be a normally distributed random variable with mean 5 and variance 9.

a State $P(X > 5)$. (1 mark)

Let Z be the standard normal random variable.

b Find the value of b such that $P(X > 7) = P(Z < b)$. (2 marks)

12 (5 marks) Let the random variable X be normally distributed with mean 2.5 and standard deviation 0.3. Let Z be defined as $Z \sim N(0, 1)$.

a Find b such that $P(X > 3.1) = P(Z < b)$. (2 marks)

b Using the fact that $P(Z < -1) = 0.16$ correct to two decimal places, find $P(X < 2.8 \mid X > 2.5)$, rounding your answer correct to two decimal places. (3 marks)

13 (6 marks) Let X be a normally distributed random variable with a mean of 72 and a standard deviation of 8. Let Z be the standard normal random variable. Use the result that $P(Z < 1) = 0.84$ correct to two decimal places, to find

a the probability that X is greater than 80 (2 marks)

b the probability that X is between 64 and 72 (2 marks)

c the probability that X is less than 64, given that it is less than 72. (2 marks)

14 (2 marks) The random variable X is normally distributed with mean 100 and standard deviation 4. If $P(X < 106) = q$, express $P(94 < X < 100)$ in terms of q.

15 © SCSA MM2018 Q2 MODIFIED (6 marks) The heights of a large group of women are normally distributed with a mean $\mu = 163$ cm and standard deviation $\sigma = 7$ cm.

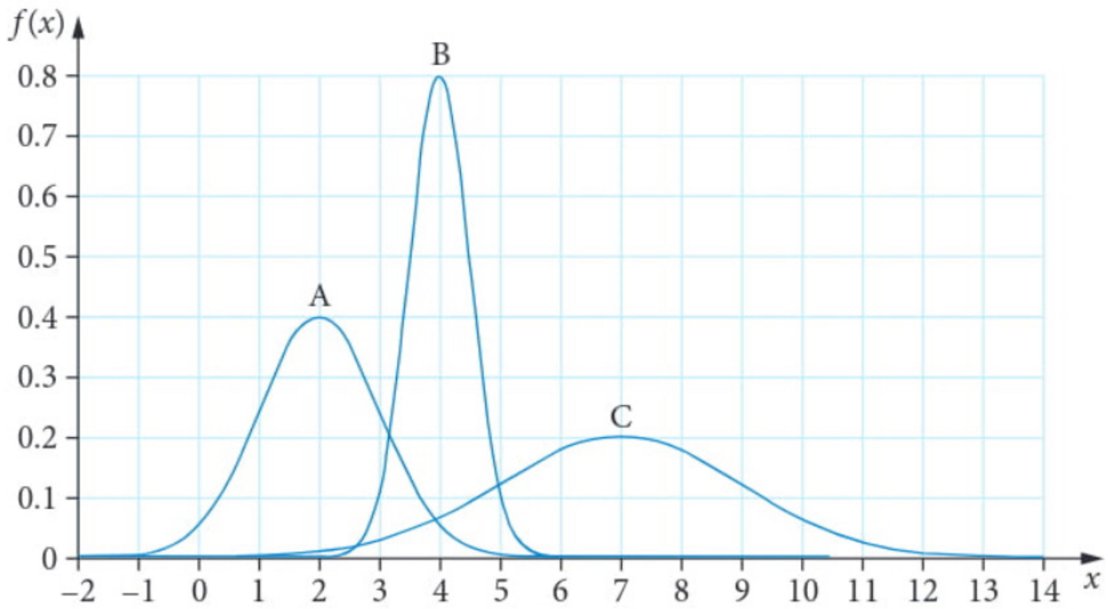
a A statistician says that almost all of the women have heights in the range 142 cm to 184 cm. Comment on the validity of her statement. Justify your answer. (2 marks)

b Approximately what percentage of women in the group has a height greater than 170 cm? (2 marks)

c Approximately 2.5% of the women are shorter than what height? (2 marks)

16 © SCSA MM2021 Q6 (7 marks)

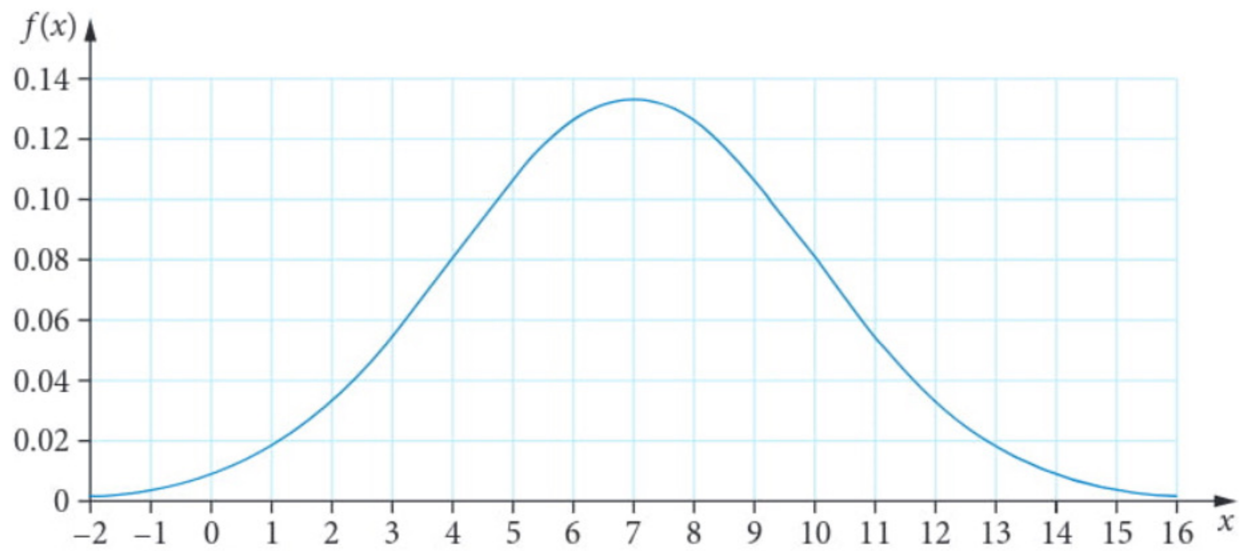
a The graphs of three normal distributions are displayed below. The distributions have been labelled A, B and C.



i What is the mean of distribution A? (1 mark)

ii Which of the distributions has the largest standard deviation? Justify your answer. (1 mark)

b A random variable X is normally distributed. The distribution of X is graphed below.



i Copy the graph and on it shade the region with area corresponding to $P(6 \le X \le 9)$. (1 mark)

ii Is $P(6 \le X \le 9) \ge 0.5$? Justify your answer. (2 marks)

c A random variable Y has probability $P(Y \ge 2) > P(Y > 2)$. Explain whether it is possible for the distribution of Y to be normal or binomial. (2 marks)

Calculator-assumed

17 (5 marks) The weights of packets of lollies are normally distributed with a mean of 200 g. It is known that 97% of these packets of lollies have a weight of more than 190 g.

a Determine the standard deviation of the distribution, correct to one decimal place. (3 marks)

b Hence, determine the probability that a randomly selected packet of lollies will have a weight between 195 g and 205 g. (2 marks)

18 © SCSA MM2016 Q18 (6 marks) The waiting times at a Perth Airport departure lounge have been found to be normally distributed. It is observed that passengers wait for less than 55 minutes, 5% of the time, while there is a 13% chance that the waiting times will be greater than 100 minutes.

a Determine the mean and standard deviation for the waiting times at Perth Airport departure lounge. (5 marks)

b Determine the probability that the waiting time will be between 75 and 90 minutes. (1 mark)

19 © SCSA MM2020 Q8 MODIFIED (7 marks) The weight, X, of chicken eggs from a farm is normally distributed with mean 60 g and standard deviation 5 g. Eggs with a weight of more than 67 g are classed as 'large'.

a What proportion of eggs from the farm are 'large'? (2 marks)

b What proportion of 'large' eggs are less than 75 g in weight? (3 marks)

c The heaviest 0.05% of eggs fetch a higher price. What is the minimum weight of these eggs? (2 marks)

20 © SCSA MM2019 Q11 (8 marks) A pizza company runs a marketing campaign based on the delivery times of its pizzas. The company claims that it will deliver a pizza in a radius of 5 km within 30 minutes of ordering or it is free. The manager estimates that the actual time, T, from order to delivery is normally distributed with mean 25 minutes and standard deviation 2 minutes.

a What is the probability that a pizza is delivered free? (1 mark)

b On a busy Saturday evening, a total of 50 pizzas are ordered. What is the probability that more than three are delivered free? (2 marks)

The company wants to reduce the proportion of pizzas that are delivered free to 0.1%.

c The manager suggests this can be achieved by increasing the advertised delivery time. What should the advertised delivery time be? (2 marks)

After some additional training the company was able to maintain the advertised delivery time as 30 minutes but reduce the proportion of pizzas delivered free to 0.1%.

d Assuming that the original mean of 25 minutes is maintained, what is the new standard deviation of delivery times? (3 marks)

21 © SCSA MM2021 Q8 (9 marks) The weights W (in grams) of carrots sold at a supermarket have been found to be normally distributed with a mean of 142.8 g and a standard deviation of 30.6 g.

a Determine the percentage of carrots sold at the supermarket that weigh more than 155 g. (2 marks)

Carrots sold at the supermarket are classified by weight, as shown in the table below.

Classification	Small	Medium	Large	Extra large
Weight W (grams)	$W \leq 110$	$110 < W \leq 155$	$155 < W \leq 210$	$W > 210$
$P(W)$		0.5131	0.3310	

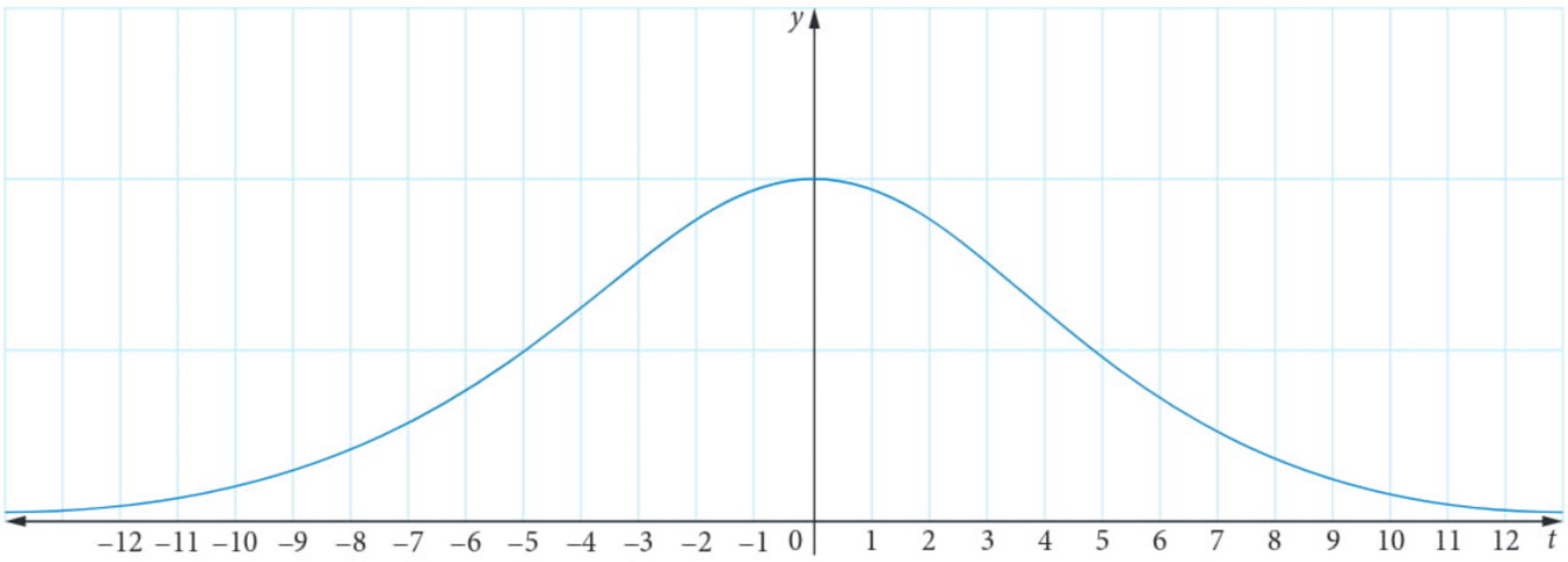
b Copy and complete the table above, providing the missing probabilities. (2 marks)

c Of the carrots being sold at the supermarket that are **not** of medium weight, what proportion is small? (2 marks)

The supermarket sells bags of mixed-weight carrots, with 12 randomly-selected carrots placed in each bag.

d If a customer purchases a bag of mixed-weight carrots, determine the probability that there will be at most two small carrots in the bag. (3 marks)

22 (12 marks) A transport company has detailed records of all its deliveries. The number of minutes a delivery is made before or after its scheduled delivery time can be modelled as a normally distributed random variable, T, with a mean of zero and a standard deviation of four minutes. A graph of the probability distribution of T is shown below.



a If $P(T \leq a) = 0.6$, find a to the nearest minute. (1 mark)

b Find the probability, correct to four decimal places, of a delivery being no later than three minutes after its scheduled delivery time, given that it arrives after its scheduled delivery time. (2 marks)

Using the model described, the transport company can make 46.48% of its deliveries over the interval $-3 \leq t \leq 2$.

c With an improvement to the delivery model, 46.48% of the transport company's deliveries can be made over the interval $-4.5 \leq t \leq 0.5$ such that the mean of T is k, while the standard deviation stays as four minutes. Find the value(s) of k, correct to one decimal place. (3 marks)

A rival transport company claims that there is a 0.85 probability that each delivery it makes will arrive on time or earlier. Assume that whether each delivery is on time or earlier is independent of other deliveries.

d Assuming that the rival company's claim is true, find the probability that on a day in which the rival company makes eight deliveries, fewer than half of them arrive on time or earlier. Give your answer correct to four decimal places. (3 marks)

e Assuming that the rival company's claim is true, consider a day in which it makes n deliveries.

i Express, in terms of n, the probability that one or more deliveries **will not** arrive on time or earlier. (1 mark)

ii Hence, or otherwise, find the minimum value of n such that there is at least a 0.95 probability that one or more deliveries **will not** arrive on time or earlier. (2 marks)

8 Chapter summary

Continuous random variables

- $P(X = k) = 0$, where k is a discrete outcome.
- The **probability density function** $f(x)$ is a piece-wise-defined function that is used to calculate probabilities.
- Probabilities can be calculated using the area under the graph of the probability density function and are found by integration, $P(a \le x \le b) = \int_a^b f(x)\,dx$, or the area formulas for triangles, rectangles or trapeziums.
- A valid probability density function has $f(x) \ge 0$ for all values of x and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
- The **cumulative distribution function** $F(x)$ is the integral of $f(x)$ such that

$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x f(t)\,dt & a \le x \le b \\ 1 & x > b \end{cases}$$

- The **expected value** (mean) of X is given by

$$E(X) = \mu = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

- The **median** $x = m$ of X is given by

$$P(X \le m) = \int_{-\infty}^{m} f(x)\,dx = 0.5$$

- The value of the pth **percentile** $(x = k)$ is determined by the integral equation

$$P(X \le k) = \frac{p}{100} \Leftrightarrow \int_{-\infty}^{k} f(x)\,dx = \frac{p}{100}$$

- The **variance** of X is given by

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx$$

or

$$\text{Var}(X) = E(X^2) - E(X)^2 = \int_a^b x^2 f(x)\,dx - \mu^2$$

- The **standard deviation** of X is given by

$$SD(X) = \sqrt{\text{Var}(X)}$$

Linear transformation $Y = aX + b$ of a continuous random variable X

- $E(aX + b) = a\,E(X) + b$
- $\text{Var}(aX + b) = a^2\,\text{Var}(X)$
- $SD(aX + b) = |a|\,SD(X)$

Uniformly distributed continuous random variable

For a **uniformly distributed continuous random variable**, X, defined over $a \le x \le b$:

- the distribution is denoted as $X \sim U[a, b]$
- the probability density function of X is defined as

$$f(x) = \begin{cases} \frac{1}{b - a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- the cumulative distribution function of X is defined as

$$f(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

- the expected value (mean) and median of X is given by

$$\mathrm{E}(X) = \frac{a+b}{2}$$

- the variance of X is given by

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

- the standard deviation of X is given by

$$\mathrm{SD}(X) = \frac{b-a}{\sqrt{12}}$$

Triangular continuous random variable

For a **triangular continuous random variable**, X, defined over $a \le x \le b$:

- the maximum value of probability density function of X occurs at the **mode** $x = c$ and has the value

$$f(c) = \frac{2}{b-a}$$

The normal distribution

For a **normally distributed continuous random variable**, X:

- the distribution is denoted as $X \sim N(\mu, \sigma^2)$
- the probability density function is given by the equation $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ and is a bell-shaped curve that is symmetrical about its mean μ, which is the same as its median and mode.

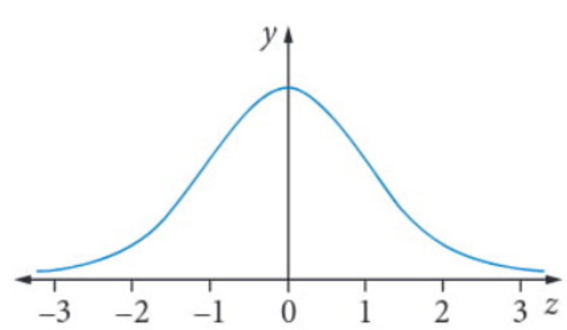
The curve satisfies the 68–95–99.7% rule such that

- $\mathrm{P}(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$
- $\mathrm{P}(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$
- $\mathrm{P}(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.997$.

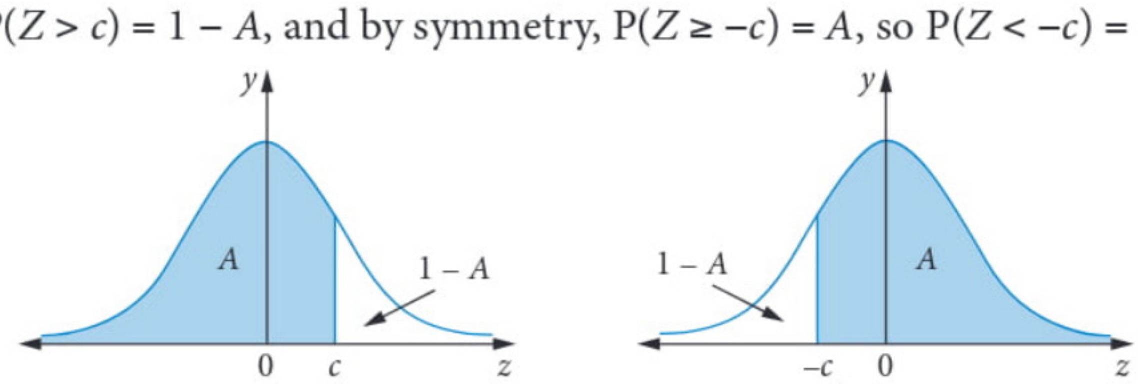
For a **standard normal random variable**, Z:

- the distribution is denoted as $Z \sim N(0, 1)$
- the probability density function is given by the equation

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z}$$



- $\mathrm{P}(Z \le c) = A$, then $\mathrm{P}(Z > c) = 1 - A$, and by symmetry, $\mathrm{P}(Z \ge -c) = A$, so $\mathrm{P}(Z < -c) = 1 - A$



- the values of Z can be obtained from $X \sim N(\mu - \sigma^2)$ using the process of **standardisation**, $Z = \dfrac{X - \mu}{\sigma}$.

Cumulative examination: Calculator-free

Total number of marks: 34 Reading time: 4 minutes Working time: 34 minutes

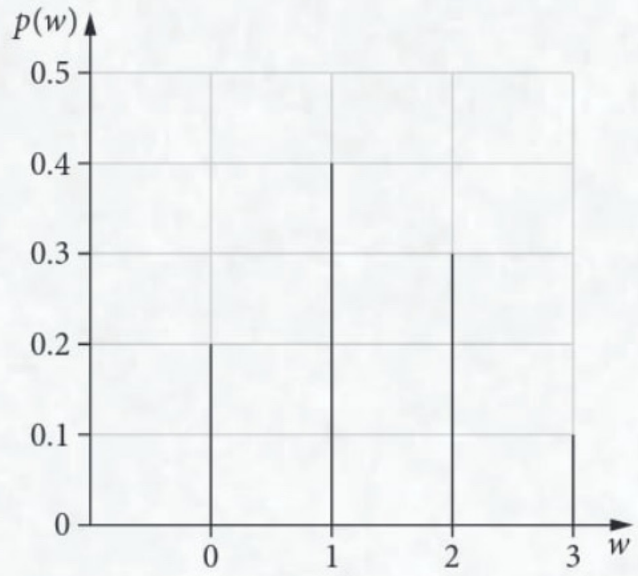
1 (4 marks) Four identical balls are numbered 1, 2, 3 and 4 and put into a box. A ball is randomly drawn from the box, and not returned to the box. A second ball is then randomly drawn from the box.

a What is the probability that the first ball drawn is numbered 4 and the second ball drawn is numbered 1? (1 mark)

b What is the probability that the sum of the numbers on the two balls is 5? (1 mark)

c Given that the sum of the numbers on the two balls is 5, what is the probability that the second ball drawn is numbered 1? (2 marks)

2 (5 marks) The graph of the distribution of the discrete random variable W is shown below.



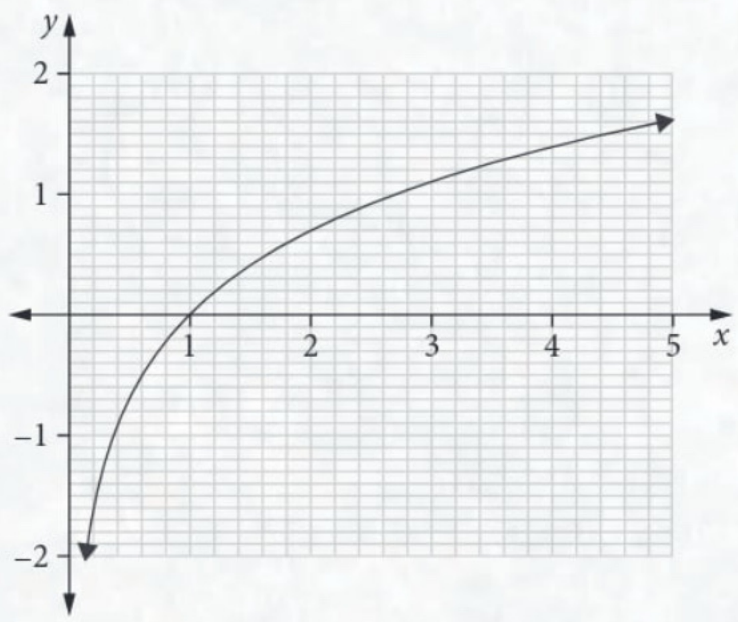
a List the probability distribution. (1 mark)

b Find

i $E(W)$ (2 marks)

ii $\text{Var}(W)$. (2 marks)

3 (3 marks) Use the graph of $y = \ln(x)$ below to find the approximate solutions to the equations.



a $\ln(x) = 1.5$ (1 mark)

b $e^{3-2x} = 0.8$ (2 marks)

4 (7 marks) Find

a $\int 2\ln(x^4)\,dx$ (2 marks)

b $\int \frac{6x+15}{x^2+5x-11}\,dx$ (2 marks)

c $\int_1^3 \frac{3}{3x+1}\,dx$. (3 marks)

5 (5 marks) A continuous random variable X has a probability density function

$$f(x) = \begin{cases} \frac{\pi}{4}\cos\left(\frac{\pi x}{4}\right) & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $\frac{d}{dx}\left(x\sin\left(\frac{\pi}{4}x\right)\right) = \frac{\pi x}{4}\cos\left(\frac{\pi x}{4}\right) + \sin\left(\frac{\pi x}{4}\right)$. (2 marks)

b Hence, determine the expected value of X. (3 marks)

6 © SCSA MM2021 Q2 (10 marks) It takes Nahyun between 15 and 40 minutes to get to school each day, depending on traffic conditions. Nahyun leaves home for school at 8:00 am each school day. Let the random variable X be the time, in minutes after 8:00 am, that Nahyun arrives at school. The probability density function of X is shown below.

a What is the name of this type of distribution? (1 mark)

b Determine:

i the values of p, q and k (2 marks)

ii the expected value of X (1 mark)

iii the probability that Nahyun arrives at school before 8:25 am. (2 marks)

Nahyun will be late for her first class if she arrives at school after 8:28 am. Otherwise, she will not be late.

c If Nahyun is not late for her first class, what is the probability that she arrives after 8:25 am? (2 marks)

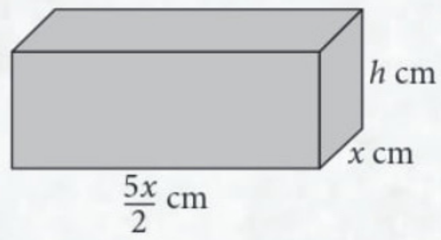
d If Nahyun only wants to be late for her first class at most 4% of the time, what time should she leave home, assuming the 15 to 40 minute travel time remains the same? (2 marks)

CUMULATIVE EXAMINATION: CALCULATOR-FREE

Cumulative examination: Calculator-assumed

Total number of marks: 56 Reading time: 6 minutes Working time: 56 minutes

1 (9 marks) A solid block in the shape of a rectangular prism has a base of width x cm. The length of the base is two-and-a-half times the width of the base. The block has a total surface area of 6480 cm^2.



a Show that if the height of the block is h cm, $h = \dfrac{6480 - 5x^2}{7x}$. (2 marks)

b The volume, $V\text{ cm}^3$, of the block is given by $V(x) = \dfrac{5x(6480 - 5x^2)}{14}$.

Given that $V(x) > 0$ and $x > 0$, find the possible values of x. (2 marks)

c Find $\dfrac{dV}{dx}$, expressing your answer in the form $\dfrac{dV}{dx} = ax^2 + b$, where a and b are real numbers. (3 marks)

d Find the exact values of x and h if the block is to have maximum volume. (2 marks)

2 © SCSA MM2020 Q15 (9 marks) A chef needs to use an oven to boil 100 mL of water in five minutes for a new experimental recipe. The temperature of the water must reach 100°C in order to boil. The temperature, T, of 100 mL of water t minutes after being placed in an oven set to T_0 can be modelled by the equation below.

$$T(t) = T_0 - 175e^{-0.07t}$$

In a preliminary experiment, the chef placed a 100 mL bowl of water into an oven that had been heated to $T_0 = 200$°C.

a What is the temperature of the water at the moment it is placed into the oven? (1 mark)

b What is the temperature of the water five minutes after being placed in the oven? (1 mark)

c What change could be made to the temperature at which the oven is set in order to achieve the five-minute boiling requirement? (2 marks)

Assume that T_0 is still 200°C.

d Determine the rate of increase in temperature of the water five minutes after being placed in the oven. Give your answer rounded to two decimal places. (2 marks)

e Explain what happens to the rate of change in the temperature of the water as time increases and how this relates to the temperature of the water. (3 marks)

3 © SCSA MM2020 Q10 (7 marks) Water flows into a bowl at a constant rate. The water level, h, measured in centimetres, increases at a rate given by

$$h'(t) = \frac{4t + 1}{2t^2 + t + 1}$$

where the time t is measured in seconds.

a Determine the rate that the water level is rising when $t = 2$ seconds. (1 mark)

b Explain why $h(t) = \ln(2t^2 + t + 1) + c$. (2 marks)

c Determine the total change in the water level over the first 2 seconds. (1 mark)

The bowl is filled when the water level reaches $\ln(56)$ cm.

d If the bowl is initially empty, determine how long it takes for the bowl to be filled. (3 marks)

4 © SCSA MM2017 Q19 (12 marks) A global financial institution transfers a large aggregate data file every evening from offices around the world to its Hong Kong head office. Once the file is received it must be processed in the company's data warehouse. The time T required to process a file is normally distributed with a mean of 90 minutes and a standard deviation of 15 minutes.

a An evening is selected at random. What is the probability that it takes more than two hours to process the file? (2 marks)

b What is the probability that the process takes more than two hours on two out of five days in a week? (3 marks)

The company is considering outsourcing the processing of the files.

c **i** A quotation for this job from an IT company is given in the table below. Copy and complete the table. (1 mark)

Job duration (minutes)	$T \leq 60$	$60 < T < 120$	$T \geq 120$
Probability			
Cost Y ($)	200	600	1200

ii What is the mean cost? (2 marks)

iii Calculate the standard deviation of the cost. (2 marks)

iv In the following year, the cost (currently $$Y$) will increase due to inflation and also the introduction of an additional fixed cost, so the new cost $$N$ is given by: $N = aY + b$. In terms of a and/or b, state the mean cost in the following year and the standard deviation of the cost in the following year. (2 marks)

5 © SCSA MM2018 Q12 (19 marks) The manager of the mail distribution centre in an organisation estimates that the weight, x (kg), of parcels that are posted is normally distributed, with mean 3 kg and standard deviation 1 kg.

a What percentage of parcels weigh more than 3.7 kg? (2 marks)

b Twenty parcels are received for posting. What is the probability that at least half of them weigh more than 3.7 kg? (3 marks)

The cost of postage, ($) y, depends on the weight of a parcel as follows:

- a cost of $5 for parcels 1 kg or less
- an additional variable cost of $1.50 for every kilogram or part thereof above 1 kg to a maximum of 4 kg
- a cost of $12 for parcels above 4 kg.

c Copy and complete the probability distribution table for Y. (4 marks)

x	≤ 1	$1 < x \leq 2$	$2 < x \leq 3$	$3 < x \leq 4$	$x > 4$
y	$5				
$P(Y = y)$					

d Calculate the mean cost of postage per parcel. (2 marks)

e Calculate the standard deviation of the cost of postage per parcel. (3 marks)

f If the cost of postage is increased by 20% and a surcharge of $1 is added for all parcels, what will be the mean and standard deviation of the new cost? (3 marks)

g Show one reason why the given normal distribution is not a good model for the weight of the parcels. (2 marks)