3.

Continuous random variables

- Distribution as a histogram
- Continuous random variables
- Probability density function (pdf)
- Uniform (or rectangular) distributions
- Non-uniform distributions
- Expected value, variance and standard deviation
- Change of scale and origin
- Cumulative distribution function
- Miscellaneous exercise three

Situation

Suppose that at a particular time the people living in Australia, and aged less than 100, had the following age distribution:

$0 \le \text{age} < 50$:	15 613 000 people.
$50 \le \text{age} < 100$:	7 517 000 people.
Total:	23 130 000 people.

1 If one of these people is selected at random what is the probability of the person being

a under 50?

b from 50 to not yet 100?

Suppose we have more detailed information about the ages of these people:

$0 \le \text{age} < 25$:	7 481 000 people.
$25 \le \text{age} < 50$:	8 132 000 people.
$50 \le \text{age} < 75$:	6 046 000 people.
$75 \le \text{age} < 100$:	1 471 000 people.
Total:	23 130 000 people.

2 If one of these people is selected at random what is the probability of the person being

a under 25?

b from 75 to not yet 100?

If one of the people who is such that $50 \le \text{age} < 100$ is chosen at random what is the probability that their age is such that

c $50 \le \text{age} < 75$?

d $75 \le \text{age} < 100$?

Even more detailed information is given below:

$0 \le \text{age} < 10$:	2 973 000 people.
$10 \le \text{age} < 20$:	2 866 000 people.
$20 \le \text{age} < 30$:	3 370 000 people.
$30 \le \text{age} < 40$:	3 210 000 people.
$40 \le \text{age} < 50$:	3 194 000 people.
$50 \le \text{age} < 60$:	2 942 000 people.
$60 \le \text{age} < 70$:	2 322 000 people.
$70 \le \text{age} < 80$:	1 371 000 people.
$80 \le \text{age} < 90$:	734 000 people.
$90 \le \text{age} < 100$:	148 000 people.
Total:	23 130 000 people.

3 If one of these people is selected at random what is the probability of the person being

a under 30?

b from 40 to not yet 100?

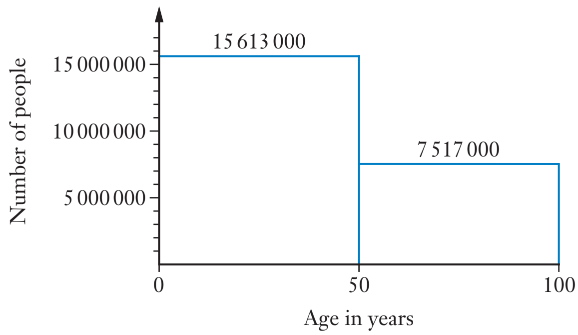
If one of the people who is less than 80 years of age is chosen at random what is the probability of the person being

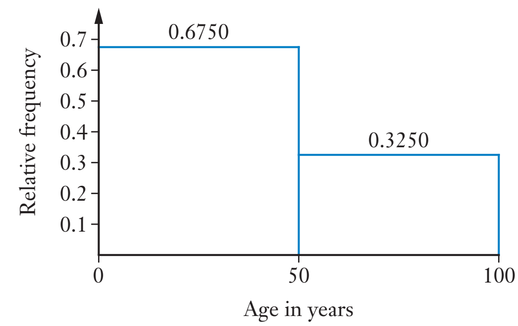
c 60 or more?

(Based on information from the Australian Bureau of Statistics.)

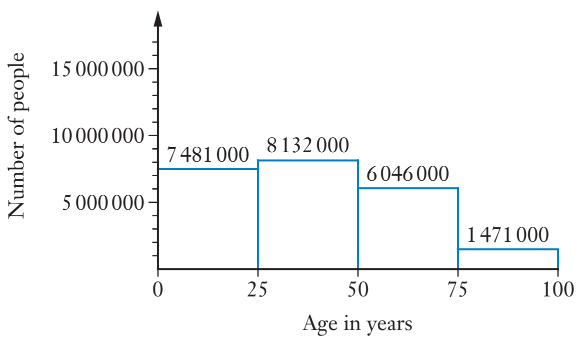
Distribution as a histogram

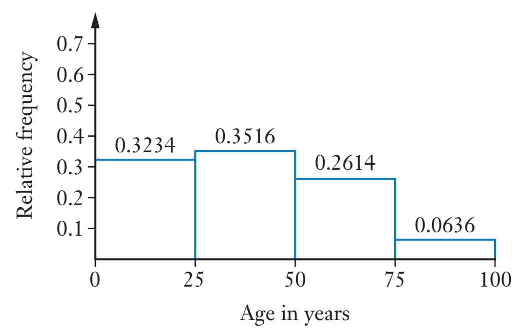
The initial information given in the situation on the previous page could be presented as a histogram showing the real data, below left, or showing the relative frequencies, below right.

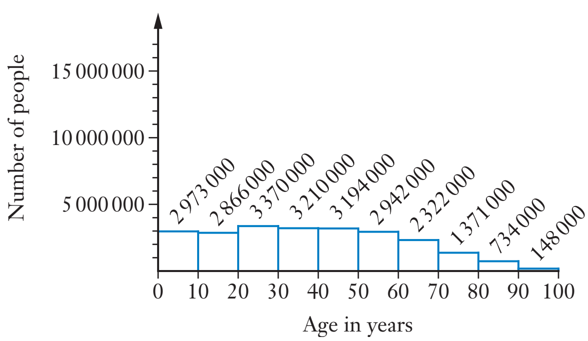


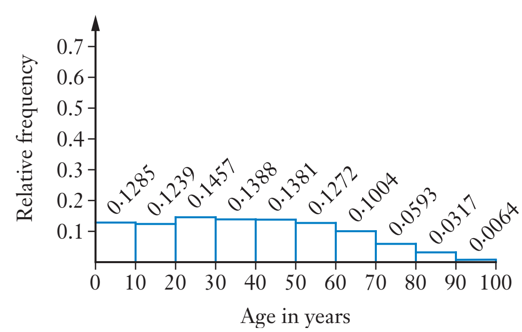


The more detailed information can also be displayed in this way, we simply have more columns:









The data forms columns on the histogram because the data was presented in 'blocks' or 'bins', e.g. under 25, 25 to 49, 50 to 74. The histogram can become smoother the more blocks it is divided into (i.e. as we reduce the 'bin width').

For each of the questions on the previous page you needed to determine the probability of something occurring by using relative frequency. The relative frequency histograms shown above give us the probability of a randomly selected individual being in a particular age range. Showing how the total probability of 1 is distributed across the possible outcomes is something you are familiar with from your work on *discrete random variables*, a concept the *Preliminary work* reminded you of. The difference here is that we have a *continuous* random variable – a person's age. Age does not have to take distinct, separate values. We may talk of someone being 18 years old but in reality this means that if their age is x years then $18 \le x < 19$.

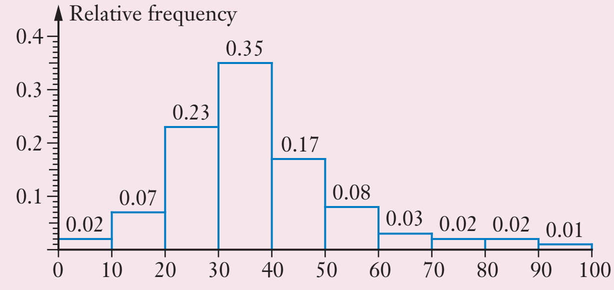
ISBN 9780170395205

EXAMPLE 1

Scientific research into a particular species of animal collects 100 of the animals, from newborn to fully grown, records various information about them, and then releases them back into the wild. One piece of recorded information, the weight of each animal, gave rise to the table shown below.

Weight (kg)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Number of animals	2	7	23	35	17	8	3	2	2	1

This table gave the following frequency histogram showing relative frequencies.



If X is the weight, in kg, of a randomly selected animal, from the 100 collected, determine:

a $P(50 \leq X < 60)$

b $P(50 \leq X < 90)$

c $P(X = 45)$

d $P(X \geq 90 \mid X > 50)$

Note: Because X is a *continuous* variable we are not restricted to particular values, as we would be with a discrete variable. Between zero and ten there are, in theory, an infinite number of values that X can take. Hence the probability of any particular value is negligible. I.e. $P(X = k) = 0$. Thus there is no difference between $P(50 \leq X \leq 60)$, $P(50 \leq X < 60)$, $P(50 < X \leq 60)$ and $P(50 < X < 60)$.

Solution

a From the table, $P(50 \leq X < 60) = \dfrac{8}{100}$ i.e. 0.08.

Or, from the graph, $P(50 \leq X < 60) = 0.08$.

b From the table, $P(50 \leq X < 90) = \dfrac{8+3+2+2}{100}$ i.e. 0.15.

Or, from the graph, $P(50 \leq X < 90) = 0.08 + 0.03 + 0.02 + 0.02 = 0.15$.

c From the note, $P(X = 45) = 0$

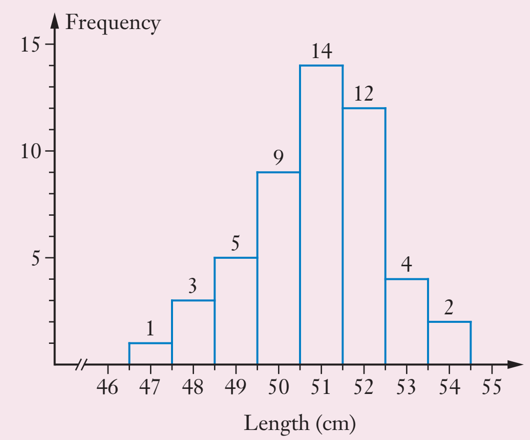
d From the table, $P(X \geq 90 \mid X > 50) = \dfrac{1}{8+3+2+2+1} = 0.0625$

Or, from the graph, $P(X \geq 90 \mid X > 50) = \dfrac{0.01}{0.08+0.03+0.02+0.02+0.01} = 0.0625$

The initial situation, involving the age distribution of Australians aged between 0 and 100, involved data based on the entire population – i.e. the ages were given for *all* people aged under 100 living in Australia (with some rounding of figures involved). Asked to determine probabilities for one item selected from the population, when we know the relevant data for the whole population, means that the probabilities will be accurate, not estimates. However, in some cases the continuous variable may be for a sample drawn from a larger population, as in example 1. We could then use the relevant frequencies of our sample to give an estimate of the probabilities for the population.

EXAMPLE 2

The lengths of 50 non-premature newborn babies born in an Australian hospital gave rise to the frequency histogram on the right:



Use the above information to suggest values for each of the following probabilities where X cm is the length of a randomly selected non-premature newborn baby born in an Australian hospital.

a $P(X \le 50.5)$

b $P(49.5 < X < 53.5)$

c $P(X > 50.5 \mid X < 53.5)$

Solution

a Of the 50 measurements that contributed to the histogram, 18 (= 1 + 3 + 5 + 9) of them involve a length less than or equal to 50.5 cm.

The given data suggests that

$$P(X \le 50.5) = \frac{18}{50} = 0.36$$

b Of the 50 measurements that contributed to the histogram, 39 (= 9 + 14 + 12 + 4) of them involve a length between 49.5 cm and 53.5 cm.

The given data suggests that $P(49.5 < X < 53.5) = 0.78$

c Given that $X < 53.5$ we need only consider the 48 lengths for which this is true.

$$P(X > 50.5 \mid X < 53.5) = \frac{P(50.5 < X < 53.5)}{P(X < 53.5)} = \frac{30}{48} \quad \text{i.e. } 0.625.$$

ISBN 9780170395205

Exercise 3A

1 The road accident statistics for a country for one year showed that for motorcyclists (drivers not passengers) in the age range fifteen to fifty-nine, 186 had died in road accidents with the distribution of the ages of these riders as shown on the right.

Age (x yrs)	Drivers killed
$15 \le x < 20$	40
$20 \le x < 25$	59
$25 \le x < 30$	29
$30 \le x < 35$	19
$35 \le x < 40$	16
$40 \le x < 45$	11
$45 \le x < 50$	8
$50 \le x < 55$	2
$55 \le x < 60$	2

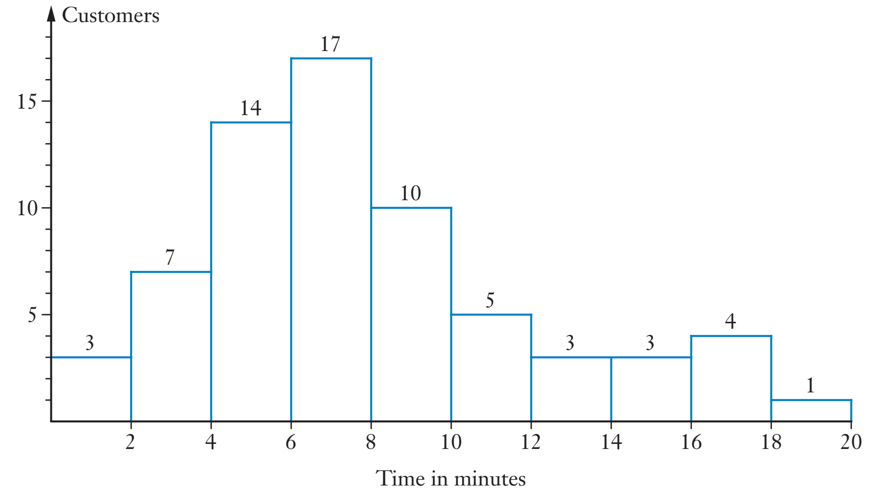
If one of these fatalities is selected at random, determine the probability that it will be for a motorcyclist aged

a less than 40 years old,

b less than 30 years old,

c at least 50 years old.

2 A pharmacy monitored the time that each customer had to wait for their prescription, from handing the prescription to the assistant, to the prescription being ready for collection. The information collected for the 67 customers requiring a prescription during one day gave rise to the following histogram.



If X minutes is the time a customer has to wait at this pharmacy for a prescription to be ready, use the information given above to suggest values for each of the following probabilities, giving your estimates as percentages to the nearest 1%.

a $P(X < 6)$

b $P(X < 4)$

c $P(X > 10)$

d $P(6 < X < 12)$

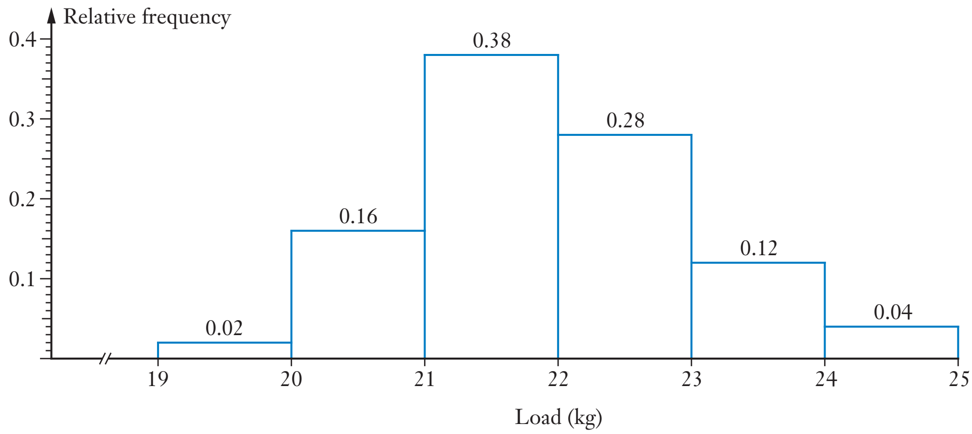
e $P(10 < X < 20)$

f $P(X > 6 | X < 10)$

iStock.com/anek_s

ISBN 9780170395205

3 To test the strength of a particular type of wire under load, samples of the wire have increasing loads attached to one end of the wire, whilst the other end is fixed in a clamp. The load that causes the wire to break is noted. The results for 50 such samples gave rise to the following relative frequency histogram.



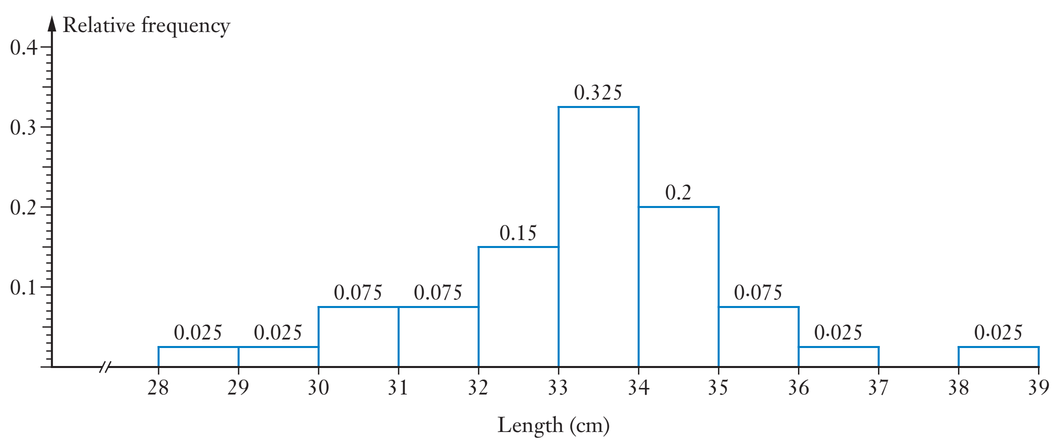
a How many of the 50 wires broke when a load between 22 kg and 24 kg was attached?

b If a random sample of wire of this type breaks when the load attached is X kg, use the above results to suggest values for each of the following probabilities.

i $P(X > 21)$ **ii** $P(X < 21)$ **iii** $P(X < 21 \mid 20 < X < 23)$

4 During research into a particular species of animal, a number of the adult male animals are caught, measured, tagged and released back into the wild.

The lengths of these animals gave rise to the following histogram of relative frequencies.



If an adult male animal of this species is captured and measured, use the above data to suggest the probability the length of the animal, L cm, is such that:

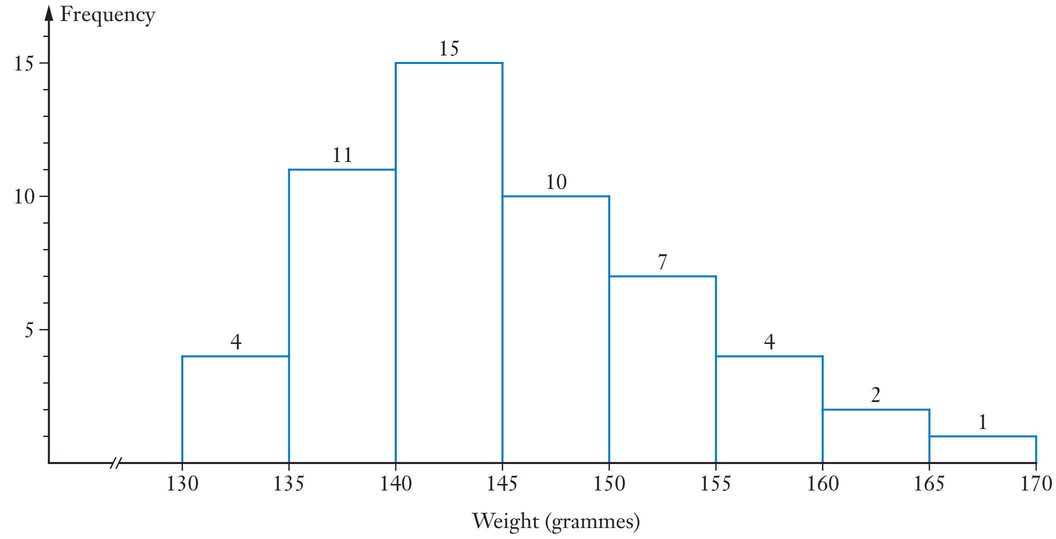
a $L \geq 33$ **b** $L > 33$

c $30 < L < 35$ **d** $L > 35$ given that $L > 33$

e $L > 33$ given that $32 < L < 36$

ISBN 9780170395205

5 A number of apples of a particular species were purchased from a supermarket. The weight of each apple was measured and noted. The distribution of weights gave rise to the following frequency histogram.



An apple of this same species is purchased from a supermarket. If the weight of this apple is W grams use the above data to suggest values for each of the following probabilities, giving your estimates as decimals, rounded to two decimal places.

a $P(W < 140)$

b $P(W \geq 140)$

c $P(140 < W < 160)$

d $P(W < 150 \mid W > 140)$

Continuous random variables

WS Discrete and continuous random variables

Each question of the previous exercise involved a *continuous* variable. The age of a motorcyclist, the length of time it takes to get a prescription ready, the load that causes a wire to break, the length of an animal and the weight of an apple are all things which can take any value, between reasonable limits appropriate to the situation. In practice the accuracy of a measurement will be limited by the measuring device we use, but if infinite accuracy were possible *any* value, between reasonable limits, would be possible.

- Discrete random variables commonly occur when we are *counting* events, for example the number of successes in a number of trials as with the binomial distribution.
- Continuous random variables commonly occur when we are *measuring* something, for example, heights, weights, times.

Suppose X is a continuous random variable that can take any value, x, in an interval. It makes no sense to talk of $P(X = x)$ because, with an infinite number of possible values, the probability of X taking any one particular value is negligible. Instead we talk of the probability of the value of X lying in some range of values, as in the previous exercise. Rather than having a probability distribution in which each of the possible values has a particular probability, as in a discrete random variable, with continuous random variables we instead talk of a **probability density function**, or **pdf**.

ISBN 9780170395205

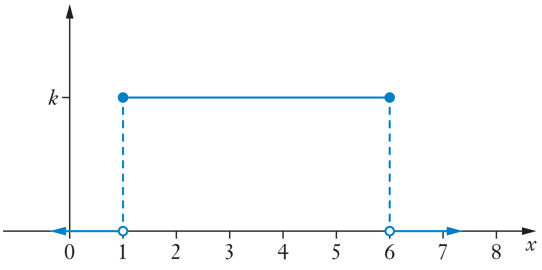
Probability density function (pdf)

WS
Probability density functions

If $f(x)$ is the probability density function for a continuous random variable, X, then the area under $f(x)$ from $x = a$ to $x = b$ gives $P(a < X < b)$.

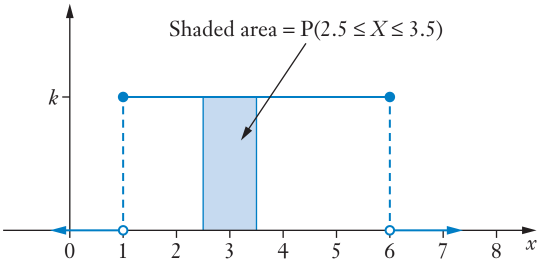
Suppose we wanted to randomly generate a number in the range 1 to 6.

In theory any number in the range is possible and we are as likely to get one in the range 1 to 2 as we are one in the range 2 to 3 or 3 to 4 etc. The distribution of probabilities is uniform across the range 1 to 6 and zero elsewhere.

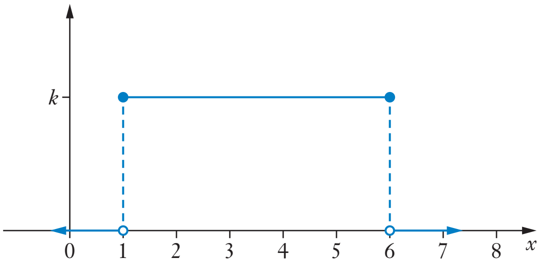


This is an example of a **uniform (or rectangular) distribution**.

For the generation of a random number in the range 1 to 6 we might be interested in $P(2.5 \le X \le 3.5)$. This can be shown as an area under the graph, as shown below.



The **probability density function**, or **pdf** of the random variable X will be the function that defines the following graph.



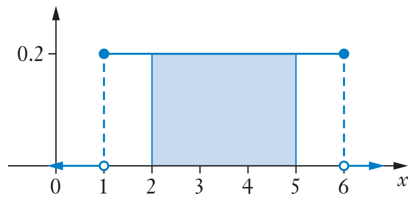
From our understanding of probability the total area under the graph must be 1. Hence for the graph above, k must be 0.2.

Hence the probability density function is given by:

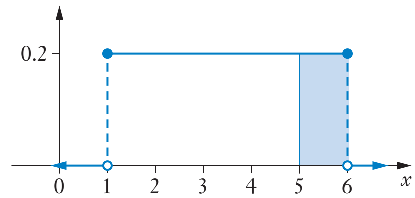
$$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \le x \le 6 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

ISBN 9780170395205

The diagrams below show $P(2 \le X \le 5)$ and $P(5 \le X \le 6)$ as areas under $f(x)$.



Shaded area $= P(2 \le X \le 5)$
$= 0.6$



Shaded area $= P(5 \le X \le 6)$
$= 0.2$

Note:

- With continuous random variables we do not determine $P(X = a)$ by evaluating $f(a)$. Instead $f(x)$ allows $P(a < X < b)$ to be determined:

  $P(a < X < b) \quad = \quad$ area under $y = f(x)$ from $x = a$ to $x = b$.

- With $P(X = a)$ being negligible it follows that $P(X \ge a) = P(X > a)$. This is consistent with the idea of the area showing probability. Whether we include a boundary line or not does not alter the area being considered.

- The graphs of $f(x)$ shown above correctly use filled and open circles to indicate where the function is and is not respectively. Thus $f(1) = 0.2$, not 0 and similarly $f(6) = 0.2$, not 0. However, whether a particular value is included or not will not alter the determination of probabilities because $P(X \ge a) = P(X > a)$. Thus when a question presents a probability density function graphically, as in the next example, the open and filled circles are often omitted.

- The previous page stated the probability density function as

  $$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \le x \le 6 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

  To avoid writing '= 0 for all other values of x' we could instead say $f(x) = 0.2$ is a probability density function defined for the interval $1 \le x \le 6$.

- The probability density function must not dip below the x-axis because that would suggest a negative probability, which is meaningless.

  Indeed for $f(x)$ to be a pdf on the interval $a < x < b$ we must have

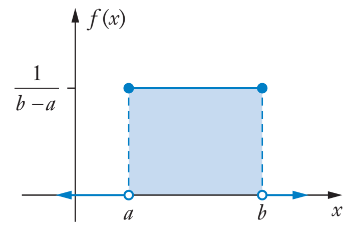
  $f(x) \ge 0$ for all x in $a < x < b$,

  and the area under $f(x)$ for $a < x < b$ must equal 1, i.e. $\int_a^b f(x)\ dx = 1$.

ISBN 9780170395205

Uniform (or rectangular) distributions

The graph below shows the probability distribution for a **uniformly distributed** continuous random variable.



For the **uniform distribution** shown the probability density function, $f(x)$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for} \quad a \le x \le b \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

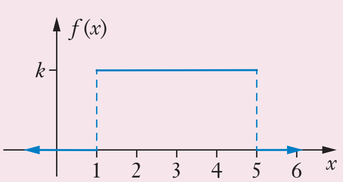
We say that the continuous random variable is **uniformly distributed on the interval $a \le x \le b$.**

Note: The interval $a \le x \le b$ is sometimes written $[a, b]$ and $a < x < b$ is sometimes written (a, b).

By symmetry, **the mean, the expected value, or the long-term average**, of the distribution will be halfway between a and b, i.e. it will equal $\dfrac{a+b}{2}$.

EXAMPLE 3

The continuous random variable X has the probability density function shown on the right.



Determine **a** k

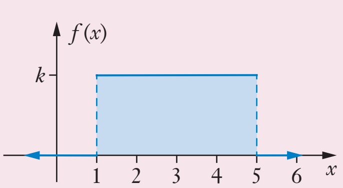
b $P(X < 3)$

c $P(2 < X < 5)$

d $P(X < 3 \mid 2 < X < 5)$

Solution

a The area of the region shaded blue in the diagram on the right must equal 1.



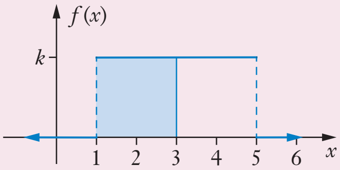
Thus $\quad 4k = 1$

$\qquad k = 0.25$

ISBN 9780170395205

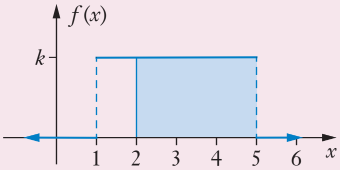
b

$$\begin{aligned} P(X < 3) &= 2k \\ &= 0.5 \end{aligned}$$



c

$$\begin{aligned} P(2 < X < 5) &= 3k \\ &= 0.75 \end{aligned}$$



d

$$\begin{aligned} P(X < 3 \mid 2 < X < 5) &= \frac{P(2 < X < 3)}{P(2 < X < 5)} \\ &= \frac{0.25}{0.75} \\ &= \frac{1}{3} \end{aligned}$$

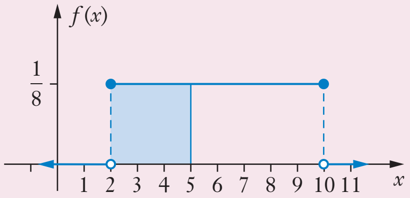
EXAMPLE 4

A continuous random variable X is uniformly distributed in the interval $2 \le X \le 10$.

Determine **a** $P(X \le 5)$ **b** $P(3 \le X \le 8)$ **c** $P(X \le 5 \mid 3 \le X \le 8)$

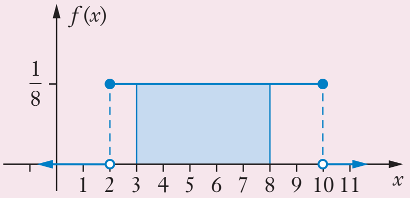
Solution

a



$$P(X \le 5) = \frac{3}{8}$$

b



$$P(3 \le X \le 8) = \frac{5}{8}$$

c

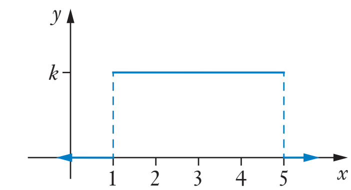
$$\begin{aligned} P(X \le 5 \mid 3 \le X \le 8) &= \frac{P(3 \le X \le 5)}{P(3 \le X \le 8)} \\ &= \frac{2}{5} \end{aligned}$$

ISBN 9780170395205

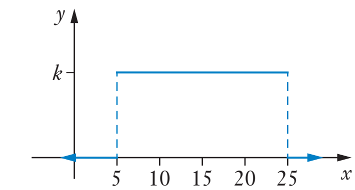
Exercise 3B

Given that each of the graphs in questions **1** to **8** show uniform probability density functions, $y = f(x)$, determine k in each case.

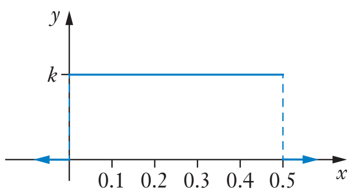
1



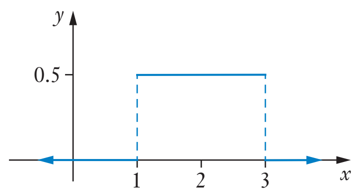
2



3

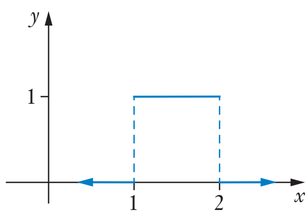


4



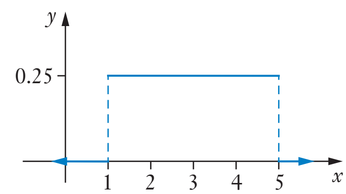
$P(X \le k) = 0.75$

5



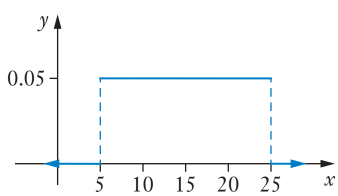
$P(X \ge k) = 0.2$

6



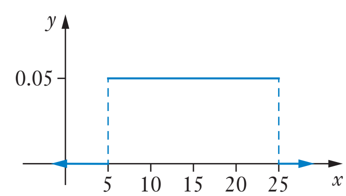
$P(X > k \mid X > 3) = 0.25$

7



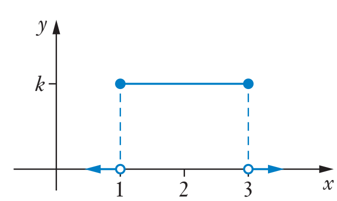
$P(X > 10 \mid X < k) = 0.5$

8



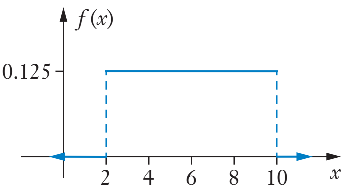
$P(X > k \mid X > 10) = 0.5$

9 Express algebraically, with the value of k determined, the uniform probability function, $f(x)$, shown graphed on the right.



ISBN 9780170395205

10 The continuous random variable X has the probability density function shown on the right.



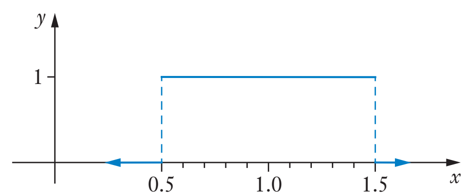
Determine **a** $P(X < 4)$

b $P(X = 4)$

c $P(X < 8)$

d $P(X > 4 | X < 8)$

11 The continuous random variable X has the probability density function shown on the right.



Determine

a $E(X)$, the long term average of X.

b $P(X > 1.2)$

c $P(X > 2)$

d $P(X < 2)$

e $P(X < 1 | X < 1.3)$

12 The continuous random variable X is uniformly distributed on the interval

$$0 \le x \le 50.$$

Find

a $E(X)$, the long term average of X.

b $P(X = 20)$

c $P(X < 20)$

d $P(X \le 20)$

e $P(X < 20 | X < 25)$

f $P(X < 25 | X < 20)$

g $P(X > 20 | X < 25)$

h $P(X > 25 | X < 20)$

13 Guided tours of a particular historic building commence every forty minutes. If we use a uniformly distributed continuous random variable X to model the time, in minutes, that a person randomly arriving at the building has to wait for the next tour to commence, show the probability density function of X graphically and find

a $P(X \le 20)$

b $P(X \ge 15)$

c $P(X \le 20 | X \ge 15)$

iStock.com/zensu

ISBN 9780170395205

Non-uniform distributions

Not all probability distributions for continuous random variables are uniform. However, whatever the shape of the distribution, if $f(x)$ is a probability density function on the interval $a < x < b$ we must have

- $f(x) \geq 0$ for all x in $a < x < b$,

and

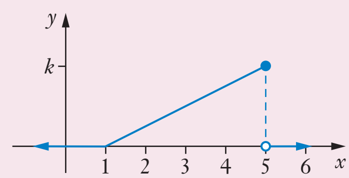
- the area under $f(x)$ for $a < x < b$ must equal 1, i.e. $\int_a^b f(x)\,dx = 1$.

EXAMPLE 5

The continuous random variable X has the probability density function shown on the right.

Determine **a** k

b $P(X \geq 3)$

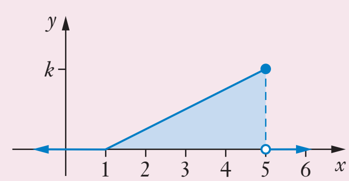


Solution

a The area of the region shaded blue in the diagram on the right must equal 1.

Thus $0.5 \times 4 \times k = 1$

$k = 0.5$



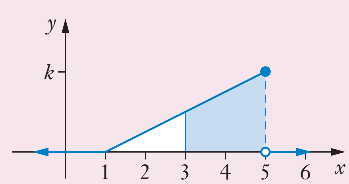
b In the diagram on the right the unshaded part of the bigger triangle is a triangle of base 2 units and height $0.5k$ units.

Unshaded area $= 0.5 \times 2 \times 0.25$ units^2

$= 0.25$ units^2

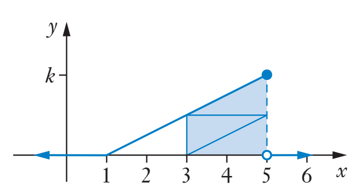
$\therefore$ Shaded area $= 0.75$ units^2

$P(X \geq 3) = 0.75$



Alternatively the answer for part **b** could be determined by:

- finding the area of the shaded trapezium directly. Shaded area $= \dfrac{0.5k + k}{2} \times 2$.
- considering the area under the curve as being made up of four triangles of equal area, three of which are shaded, as shown on the right.
- evaluating $\displaystyle\int_3^5 \left(\frac{1}{8}x - \frac{1}{8}\right) dx$.



ISBN 9780170395205

EXAMPLE 6

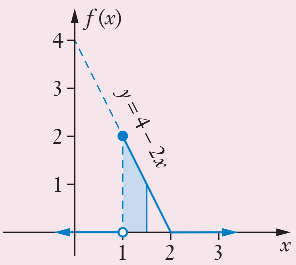
The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} 4 - 2x & \text{for} \quad 1 \le x \le 2 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Determine $P(X < 1.5)$.

Solution

The probability density function is shown on the right, with the blue shaded area representing $P(X < 1.5)$.



$$\therefore \quad P(X < 1.5) = \frac{2+1}{2} \times 0.5$$
$$= 0.75$$

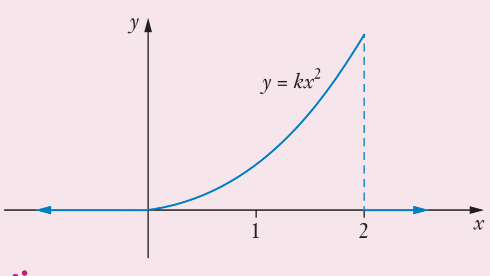
Alternatively, using calculus:

$$P(X < 1.5) = \int_1^{1.5} (4 - 2x)\, dx$$
$$= \left[4x - x^2\right]_1^{1.5}$$
$$= (6 - 2.25) - (4 - 1)$$
$$= 0.75$$

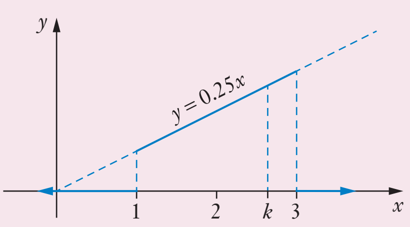
EXAMPLE 7

Each of the diagrams below show probability density functions. Use calculus to determine the value of k in each case.

a



b



$P(1 \le X \le k) = 0.75$

Solution

a

$$\int_0^2 kx^2\, dx = 1$$
$$\therefore \quad \left[\frac{kx^3}{3}\right]_0^2 = 1$$
$$\frac{8k}{3} = 1$$
$$k = \frac{3}{8}$$

b

$$\int_1^k 0.25x\, dx = 0.75$$
$$\therefore \quad \left[\frac{x^2}{8}\right]_1^k = 0.75$$
$$\frac{k^2}{8} - \frac{1}{8} = 0.75$$
$$k^2 = 7$$
$$k = \sqrt{7}$$

(Negative solution not applicable.)

Repeat part **b** of this example without using calculus.

EXAMPLE 8

Let us suppose that the continuous random variable X is the time in minutes between an event occurring and it next occurring, and that X has the probability density function:

$$f(x) = \begin{cases} 0.5e^{-0.5x} & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x \le 0. \end{cases}$$

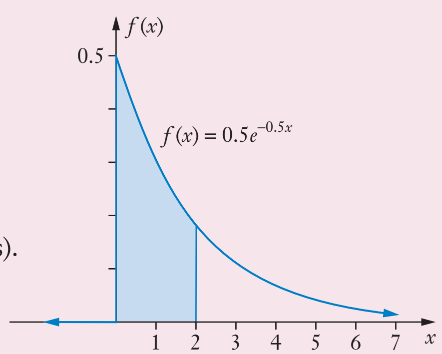
Determine **a** $P(X \le 2)$,

b $P(5 \le X \le 10)$,

c the value of d for which $P(X \le d) = 0.5$.

Solution

a $\quad P(X \le 2) = \displaystyle\int_0^2 0.5e^{-0.5x}\, dx$



Either by calculator, or algebraically, as shown below,

$$\begin{aligned} \int_0^2 0.5e^{-0.5x}\, dx &= \left[-e^{-0.5x}\right]_0^2 \\ &= -e^{-1} + e^0 \\ &= 0.6321 \text{ (correct to 4 decimal places).} \end{aligned}$$

b $\quad$ $$\begin{aligned} P(5 \le X \le 10) &= \int_5^{10} 0.5e^{-0.5x}\, dx \\ &= 0.0753 \text{ (correct to 4 decimal places)} \end{aligned}$$

c $\quad$ If $P(X \le d) = 0.5$ then $\displaystyle\int_0^d 0.5e^{-0.5x}\, dx = 0.5$

Solve algebraically, as shown below, or by calculator.

$$\begin{aligned} \therefore \quad \left[-e^{-0.5x}\right]_0^d &= 0.5 \\ -e^{-0.5d} + 1 &= 0.5 \\ 0.5 &= e^{-0.5d} \end{aligned}$$

Taking natural logs of both sides $\quad \ln 0.5 = -0.5d$

Thus $\quad d = 1.386$ (correct to 3 decimal places)

ISBN 9780170395205

Note: Probability density functions of the form

$$f(x) = \begin{cases} ke^{-kx} & \text{for} \quad x > 0 \\ 0 & \quad\quad\; \text{elsewhere,} \end{cases}$$

as in the previous example, are typically involved when the random variable is the time between an event occurring and it next occurring (the interarrival time).

For functions of this type it follows that for $0 < x < \infty$ and $k > 0$, then $f(x) > 0$ and, as the reader should confirm by performing the integration,

$$\int_0^\infty f(x)\,dx = 1.$$

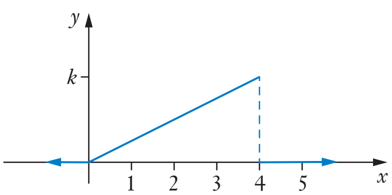
Exercise 3C

Given that each of the graphs in questions **1** to **16** show probability density functions,

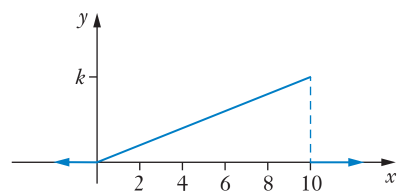
$$y = f(x),$$

determine k in each case.

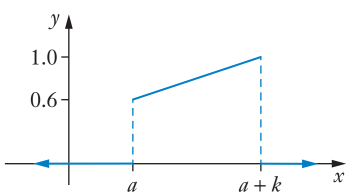
1



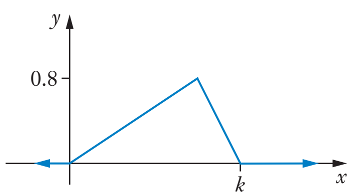
2



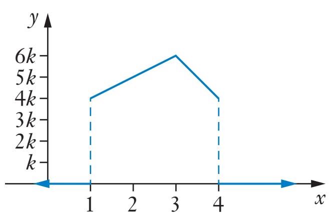
3



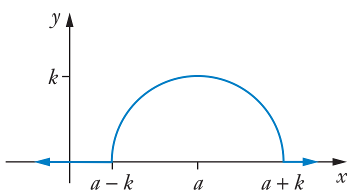
4



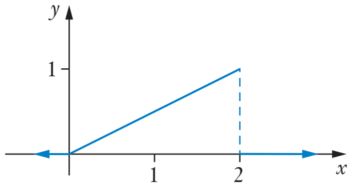
5



6

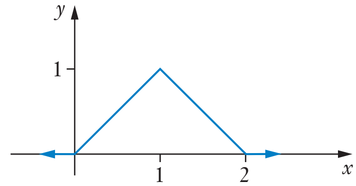


7



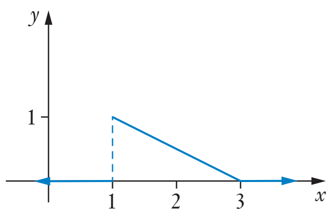
$P(X \leq k) = 0.0625$

8



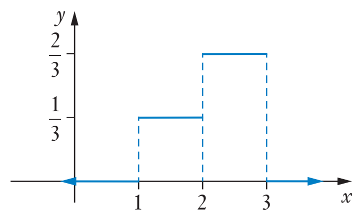
$P(X > k) = 0.02$

9



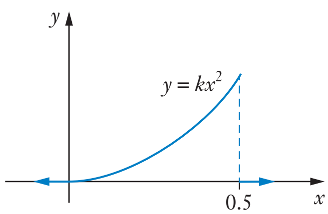
$P(X \leq k) = 0.51$

10

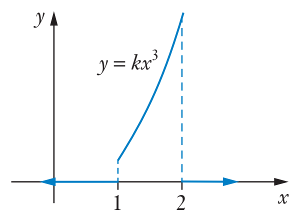


$P(X \leq k) = \frac{2}{3}$

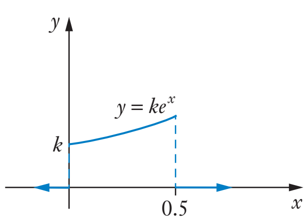
11



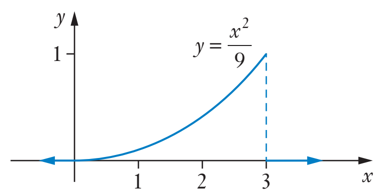
12



13

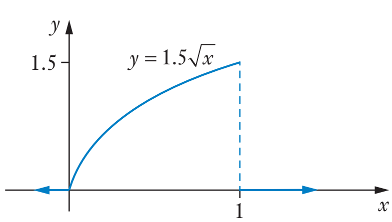


14



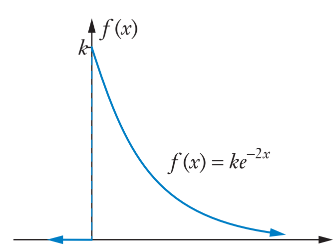
$P(X < k) = 0.512$

15



$P(X \leq k) = 0.125$

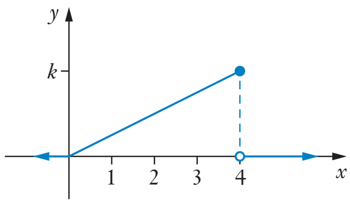
16



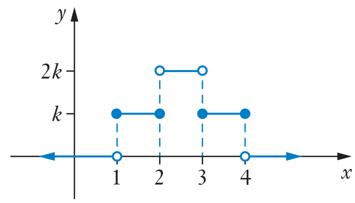
ISBN 9780170395205

17 For each of the probability density functions shown graphed below, write the probability density function, $f(x)$. (In each case k should be determined.)

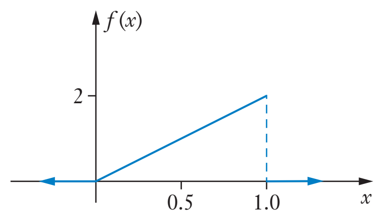
a



b



18 The continuous random variable X has the probability density function shown on the right.



Determine

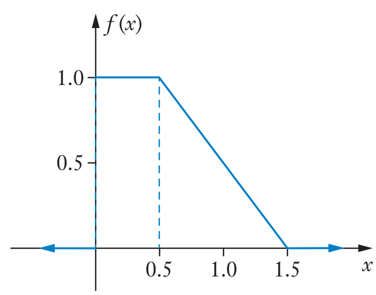
a $P(X < 0.5)$

b $P(X > 0.75)$

c $P(X < 2)$

d $P(X > 0.5 \mid X < 0.75)$

19 The continuous random variable X has the probability density function shown on the right.



Determine

a $P(X < 0.5)$

b $P(X > 1)$

c $P(X < 1)$

d $P(X > 0.5 \mid X < 1)$

20 For $0 \le x \le 5$, only one of the following functions could be a probability density function. Decide which function it is and explain why the other two cannot be probability density functions for $0 \le x \le 5$.

$f(x) = 0.5 - 0.08x \qquad g(x) = 0.5 - 0.12x \qquad h(x) = 0.4 - 0.08x$

21 A continuous random variable, X, has probability density function

$$f(x) = \begin{cases} 0.08(x+2) & \text{for } -2 \le x \le 3 \\ 0 & \text{for all other values of } x. \end{cases}$$

Determine **a** $P(X \ge 0)$ **b** $P(1 \le X \le 2)$ **c** $P(X \le 2 \mid X \ge 1)$

22 A continuous random variable, X, has probability density function

$$f(x) = \begin{cases} kx & \text{for } 0 \le x \le 5 \\ 0 & \text{for all other values of } x. \end{cases}$$

Determine **a** k **b** $P(X \le 4)$

c $P(2 \le X \le 4)$ **d** $P(X \ge 2 \mid X \le 4)$

ISBN 9780170395205

23 A continuous random variable, X, has pdf: $f(x) = \dfrac{3}{2x^2}$ defined for $[1, 3]$.

a Confirm that $\displaystyle\int_1^3 \frac{3}{2x^2}\,dx = 1.$

Determine **b** $P(X \geq 2)$ **c** $P(2 \leq X \leq 2.5)$ **d** $P(X \leq 2.5 \mid X \geq 2)$

24 **a** If
$$f(x) = \begin{cases} x^2 + kx & \text{for} \quad 1 \leq x \leq 4 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

and $\displaystyle\int_1^4 f(x)\,dx = 1$, determine k.

b For the value of k determined in part **a** could $f(x)$ represent a probability density function? (Explain your answer.)

25 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} k(1-x)(x-3) & \text{for} \quad 1 \leq x \leq 3 \\ 0 & \text{for} \quad \text{for all other values of } x \end{cases}$$

Determine **a** k,

b $P(X \leq 2)$,

c $P(X \leq 2.5)$, giving your answer correct to 2 decimal places,

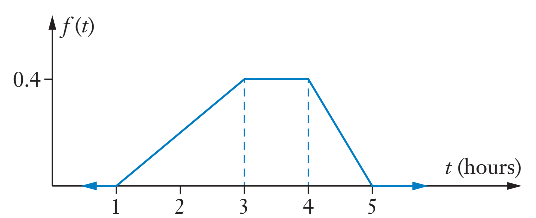
d q, correct to two decimal places, given that $P(X \geq q) = 0.6$.

26 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} \dfrac{a + bx - x^2}{9} & \text{for} \quad 1 \leq x \leq 4 \\ 0 & \text{for} \quad \text{all other values of } x \end{cases}$$

If $P(X \leq 2) = \dfrac{5}{27}$ determine a and b.

27 The time a motor mechanic may take to carry out a particular repair can depend upon a number of things. The skill and experience of the mechanic, the age, make and condition of the vehicle are just some of the factors that could influence the time taken.



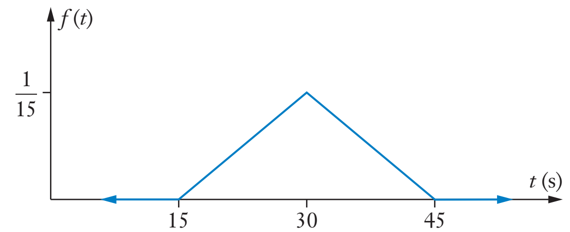
Suppose the time taken to complete a certain repair is between 1 and 5 hours with probability density function, $f(t)$, shown on the right.

a Find the probability that the repair takes less than 3 hours to complete.

b Given that the repair took more than 3 hours to complete what is the probability that it took less than 4 hours to complete?

ISBN 9780170395205

28 Suppose that the time, in seconds, that a person actually records when asked to estimate a time of 30 seconds can be represented by a random variable with the probability density function, $f(t)$, shown on the right.



Find the probability that a person set this task will

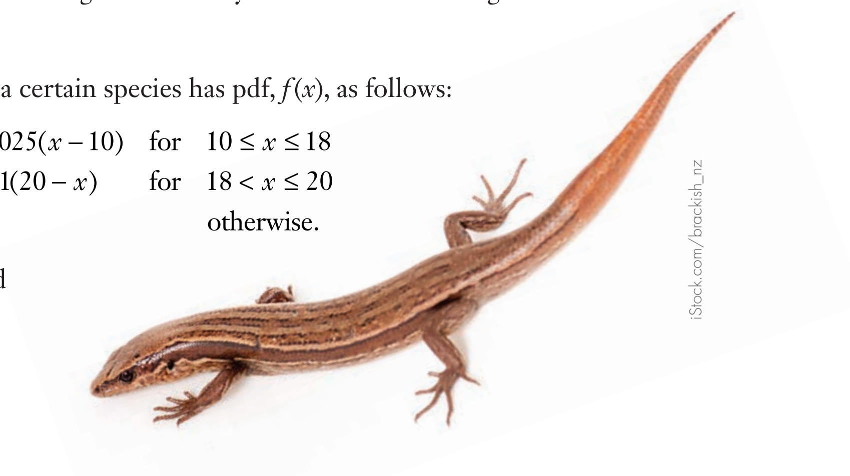
a record a time that is less than 25 seconds,

b record a time that is within 5 seconds of 30 seconds,

c record a time that is less than 40 seconds given that they record a time that is greater than 30 seconds.

29 The length, X cm, of an adult lizard of a certain species has pdf, $f(x)$, as follows:

$$f(x) = \begin{cases} 0.025(x-10) & \text{for } 10 \le x \le 18 \\ 0.1(20-x) & \text{for } 18 < x \le 20 \\ 0 & \text{otherwise.} \end{cases}$$



iStock.com/brackish_nz

Find the probability that an adult lizard of this species has a length that is

a less than 18 cm,

b greater than 14 cm,

c less than 19 cm.

30 Suppose that the number of years a particular make of washing machine lasts before its first breakdown is a continuous random variable, X, with pdf:

$$f(x) = 0.2e^{-0.2x} \qquad \text{for } (0, \infty).$$

a Determine the probability that a washing machine of this make will last more than eight years before its first breakdown. (Correct to four decimal places.)

b (Hint: Binomial distribution.) If we had six washing machines of this make, determine the probability that exactly two will last more than eight years before they experience their first breakdown, the other four experiencing their first breakdown within eight years.

Shutterstock.com/wavebreakmedia

ISBN 9780170395205

Expected value, variance and standard deviation

The *Preliminary work* reminded us of the fact that if the discrete random variable, X, has possible values x_i, with $P(X = x_i) = p_i$ then $E(X)$, the expected or long term mean value is given by:

$$E(X) = \Sigma(x_i\ p_i)$$

the summation being carried out over all of the possible values x_i.

Further, if we use the Greek letter, μ, to represent $E(X)$ then the **variance**, $\text{Var}(X)$ is given by:

$$\text{Var}(X) = \Sigma[p_i(x_i - \mu)^2]$$

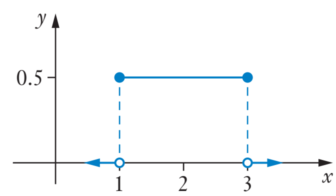
The **standard deviation** is then the square root of the variance.

For a continuous random variable X, with probability density function $f(x)$, the corresponding statements are:

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

$$\text{Var}(X) = \int_{-\infty}^{\infty} \left[f(x)(x-\mu)^2 \right] dx$$

Thus, for the pdf shown on the right,



$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\ &= \int_1^3 (x \times 0.5)\, dx \\ &= \left[\frac{x^2}{4} \right]_1^3 \\ &= \frac{9}{4} - \frac{1}{4} \\ &= 2 \quad \text{(As we would expect.)} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} \left[f(x)(x-\mu)^2 \right] dx \\ &= \int_1^3 \left[0.5(x-2)^2 \right] dx \\ &= \frac{1}{3} \end{aligned}$$

$$\int_1^3 0.5x\, dx$$

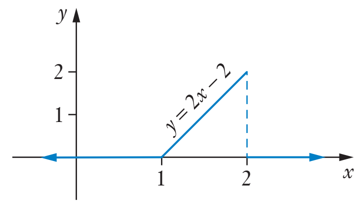
$$2$$

$$\int_1^3 0.5(x-2)^2\, dx$$

$$\frac{1}{3}$$

ISBN 9780170395205

Similarly, for the pdf shown on the right:



$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\
&= \int_1^2 x(2x-2)\, dx \\
&= \left[\frac{2x^3}{3} - x^2\right]_1^2 \\
&= \left(\frac{16}{3} - 4\right) - \left(\frac{2}{3} - 1\right) \\
&= \frac{5}{3}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} \left[f(x)(x-\mu)^2\right] dx \\
&= \int_1^2 \left[2(x-1)\left(x - \frac{5}{3}\right)^2\right] dx \\
&= \frac{1}{18}
\end{aligned}$$

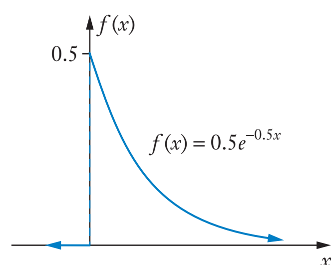
$\int_1^2 x(2x-2)\,dx$

$\frac{5}{3}$

$\int_1^2 2(x-1)(x-\frac{5}{3})^2\,dx$

$\frac{1}{18}$

For the pdf shown,



carrying out appropriate integrations on a calculator,

$$\begin{aligned}
E(X) &= 2 \\
\mathrm{Var}(X) &= 4 \\
\mathrm{SD}(X) &= 2
\end{aligned}$$

$\int_0^{\infty} 0.5xe^{-0.5x}\,dx$

2

$\int_0^{\infty} 0.5(x-2)^2 e^{-0.5x}\,dx$

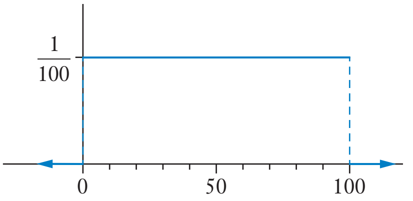
4

Change of scale and origin

Suppose that the temperature of something, recorded in °C, is a continuous random variable X.

Further suppose that X is uniformly distributed between 0 and 100.

The probability density function for X will be as shown on the right.



By symmetry, or by calculus (see right), $\mathrm{E}(X) = 50$.

$$\int_0^{100} \left(\frac{1}{100}x\right) dx = 50.$$

By calculus (see right), $\quad \mathrm{Var}(X) = \dfrac{2500}{3}.$

$$\int_0^{100} \left(\frac{1}{100}(x-50)^2\right) dx = \frac{2500}{3}.$$

Hence $\quad \mathrm{SD}(X) = \dfrac{50}{\sqrt{3}} = \dfrac{50\sqrt{3}}{3}.$

Suppose instead that the temperatures were measured in °F (degrees Fahrenheit).

To change a temperature in °C to the equivalent temperature in °F we multiply by 1.8 and add 32. I.e. °F = 1.8 × °C + 32

Hence, 0°C = 32°F
and 100°C = 212°F.

We would now have a uniform distribution (Y) involving temperatures from 32°F to 212 °F.

By symmetry, or by calculus (see right), $\mathrm{E}(Y) = 122$.

$$\int_{32}^{212} \left(\frac{1}{180}x\right) dx = 122.$$

By calculus (see right), $\quad \mathrm{Var}(Y) = 2700.$

$$\int_{32}^{212} \left(\frac{1}{180}(x-122)^2\right) dx = 2700.$$

Hence $\quad \mathrm{SD}(Y) = 30\sqrt{3}$

Notice that $\quad 122 = 1.8 \times 50 + 32$

I.e. $\quad \mathrm{E}(Y) = 1.8 \times \mathrm{E}(X) + 32$

And $\quad 30\sqrt{3} = 1.8 \times \dfrac{50\sqrt{3}}{3}$

I.e. $\quad \mathrm{SD}(Y) = 1.8 \times \mathrm{SD}(X)$

This should not really be any surprise because the *Preliminary work* section reminded us of the effect of *changes of scale and origin*.

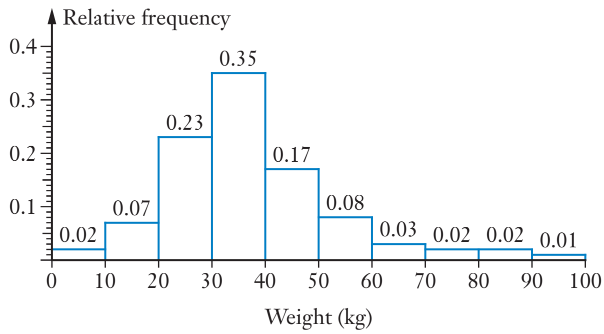
If the random variable X has mean μ and standard deviation σ (and variance σ^2) then the random variable $aX + b$ will have mean $a\mu + b$ and standard deviation $|a|\sigma$ (and variance $a^2\sigma^2$).

ISBN 9780170395205

Cumulative distribution function

Let us again consider the weights of the 100 animals of a particular species that we considered in Example 1 on page 45:

Weight (kg)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Number of animals	2	7	23	35	17	8	3	2	2	1
Relative frequency	0.02	0.07	0.23	0.35	0.17	0.08	0.03	0.02	0.02	0.01



We could present the information in terms of the 'running totals' or *cumulative* frequencies:

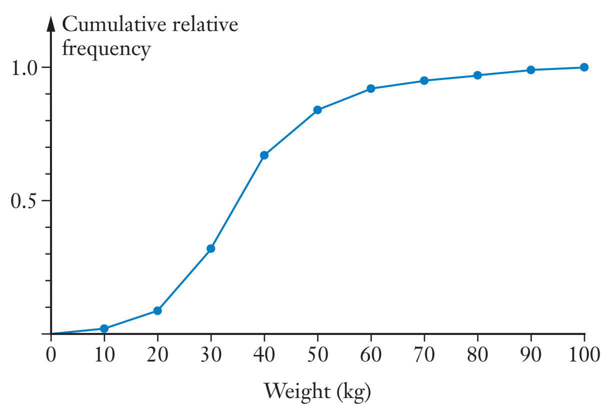
Weight (kg)	≤ 10	≤ 20	≤ 30	≤ 40	≤ 50	≤ 60	≤ 70	≤ 80	≤ 90	≤ 100
Number of animals	2	9	32	67	84	92	95	97	99	100
Relative frequency	0.02	0.09	0.32	0.67	0.84	0.92	0.95	0.97	0.99	1

As we would expect for cumulative frequencies:

- the numbers of animals increase (or could stay the same) as we move from one cell to the next cell on the right,
- the final frequency figure is 100, the total number of animals involved,

and
- the final relative frequency figure is 1.

Showing the cumulative relative frequencies as a line graph:



(A graph showing cumulative frequencies in this way is also known as an **ogive**.)

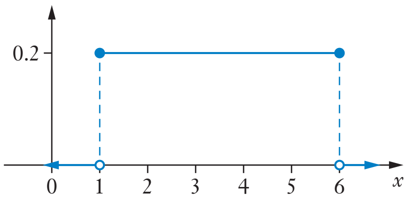
ISBN 9780170395205

Given a probability density function, pdf, for a random variable, X, we can similarly create a cumulative distribution function, **cdf**.

The cumulative distribution function, $f(x)$, will be such that $f(k)$ gives the probability that X will take a value less than or equal to k.

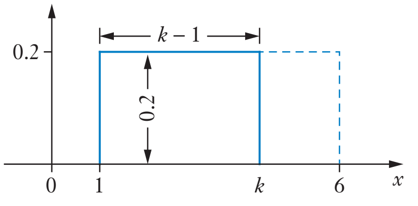
Consider the following uniform pdf:

$$f(x) = \begin{cases} 0.2 & \text{for} \quad 1 \le x \le 6 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$



For $1 \le k \le 6$,
$$\begin{aligned} P(X \le k) &= \int_1^k 0.2 \, dx \\ &= \left[0.2x\right]_1^k \\ &= 0.2(k-1) \end{aligned}$$

Note that the same formula could be obtained without integration, by considering the area of a rectangle with a base of $(k - 1)$ and height 0.2.



Hence the cumulative distribution function will be

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ 0.2(x-1) & \text{for} \quad 1 \le x \le 6 \\ 1 & \text{for} \quad x > 6 \end{cases}$$

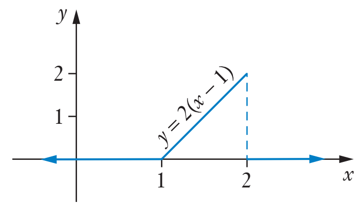
If we want to determine $P(3 \le X \le 5)$, we could then proceed as follows.

$$\begin{aligned} P(3 \le X \le 5) &= P(X \le 5) - P(X < 3) \\ &= 0.2(5-1) - 0.2(3-1) \\ &= 0.8 - 0.4 \\ &= 0.4 \end{aligned}$$

Of course this answer is exactly the same as would be obtained by considering the appropriate rectangular area, or by evaluating $\int_3^5 0.2 \, dx$. The cumulative distribution function has simply 'done the integration for us'.

ISBN 9780170395205

Consider the triangular pdf shown on the right.



For $1 \le k \le 2$,

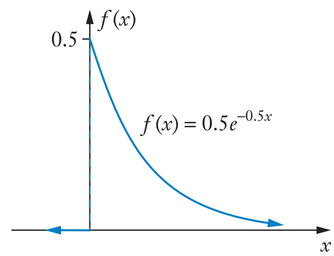
$$\begin{aligned} P(X \le k) &= \int_1^k 2(x-1)\,dx \\ &= \left[(x-1)^2\right]_1^k \\ &= (k-1)^2 \end{aligned}$$

Hence

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x \le 1 \\ (x-1)^2 & \text{for} \quad 1 < x \le 2 \\ 1 & \text{for} \quad x > 2 \end{cases}$$

(The placement of the '= part of the inequality' could differ from that shown above.)

Consider the exponential pdf shown on the right.



For $k \ge 0$,

$$\begin{aligned} P(X \le k) &= \int_0^k 0.5e^{-0.5x}\,dx \\ &= \left[-e^{-0.5x}\right]_0^k \\ &= 1 - e^{-0.5k} \end{aligned}$$

Hence the cumulative distribution function will be:

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-0.5x} & \text{for} \quad x \ge 0 \end{cases}$$

Using this function to determine $P(5 \le X \le 10)$ as an example,

$$\begin{aligned} P(5 \le X \le 10) &= P(X \le 10) \quad - \quad P(X < 5) \\ &= (1 - e^{-5}) \quad - \quad (1 - e^{-2.5}) \\ &= 0.0753 \quad \text{(correct to 4 decimal places)} \end{aligned}$$

(As obtained by determining $\int_5^{10} 0.5e^{-0.5x}\,dx$ in example 8 earlier in this chapter on page 58.)

Thus whilst a pdf allows probabilities to be determine using integration, a cdf allows probabilities to be determined by direct substitution into a formula.

If the continuous random variable X has probability density function $f(x)$ we can define the cumulative distribution function of X more formally as

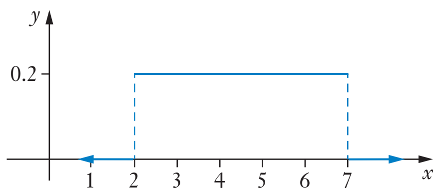
$$\int_{-\infty}^{x} f(t)\,dt$$

Exercise 3D

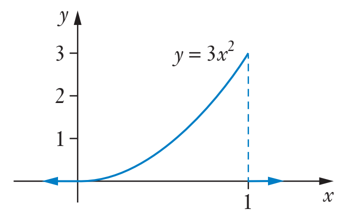
Expected value, variance and standard deviation.
(Evaluate appropriate definite integrals using your calculator if you wish.)

Use calculus to determine the mean (expected value) and variance of each of the pdfs shown in questions **1** to **4**.

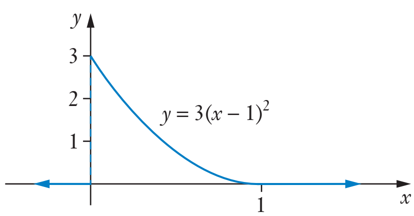
1



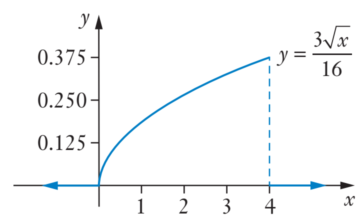
2



3



4



5 Jennifer spends X minutes in the shower, with X having the probability density function $f(x)$ defined as follows,

$$f(x) = \begin{cases} \dfrac{12-x}{50} & \text{for} \quad 2 \le x \le 12 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Find **a** $E(X)$, the expected value of X,

b the standard deviation of X.

6 The distance, in metres, between consecutive defects in a piece of wire is a continuous random variable X, with probability density function:

$$f(x) = \begin{cases} 0.01e^{-0.01x} & \text{for} \quad x > 0 \\ 0 & \quad\quad\;\; \text{elsewhere.} \end{cases}$$

Performing integrations using your calculator, determine the mean distance between defects.

7 Let us suppose that the number of hours of homework Sandy does each day is a continuous random variable X, with probability density function, $f(x)$, given by

$$f(x) = \begin{cases} \dfrac{3(6x - x^2 - 5)}{32} & \text{for} \quad 1 \le x \le 5 \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

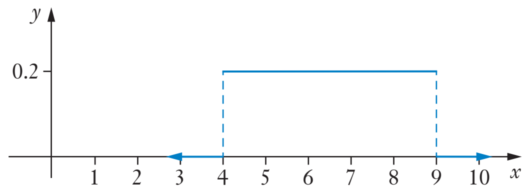
Find the mean, variance and standard deviation of X.

ISBN 9780170395205

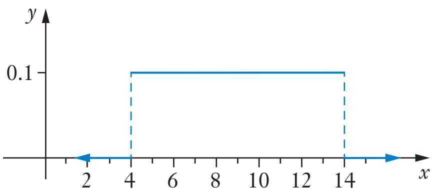
Change of scale and origin

Use your answers to question **1** to determine the mean and variance of the pdfs shown in questions **8** and **9**.

8



9



10 If a continuous random variable X has an expected value of 12 and a standard deviation of 3, find the expected value and the standard deviation of the continuous random variable Y in each of the following situations

a $Y = 3X$ **b** $Y = X + 3$ **c** $Y = 2X + 5$

11 If a continuous random variable X has an expected value of 20 and a standard deviation of 4, find the expected value and the standard deviation of the continuous random variable Z in each of the following situations

a $Z = 5X + 2$ **b** $Z = 2X + 5$ **c** $Z = 3X + 4$

12 The random variable X involves temperatures measured in degrees Celsius.

X has mean 48 and variance 16.

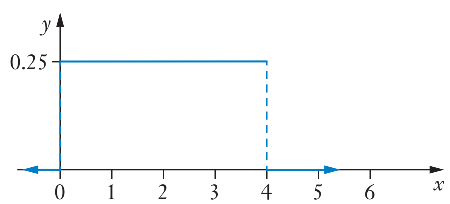
If the random variable Y involves the same temperature distribution as X but with the temperatures changed to degrees Fahrenheit, what will be the mean, variance and standard deviation of Y? (Note, °F = 1.8 × °C + 32.)

13 How would changing a random variable involving lengths measured in centimetres to the same lengths measured in metres alter the mean and standard deviation?

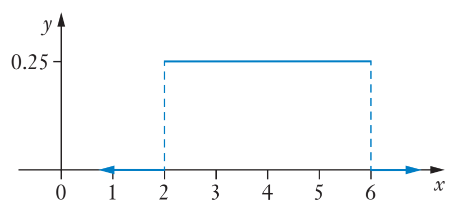
Cumulative distribution function

Express each of the following probability density functions as cumulative distribution functions.

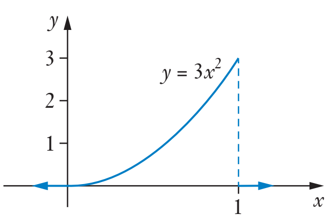
14



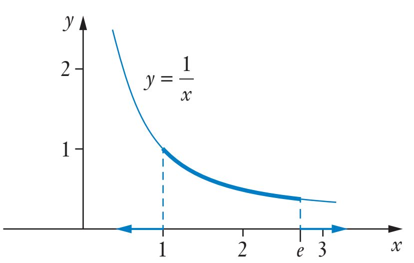
15



16

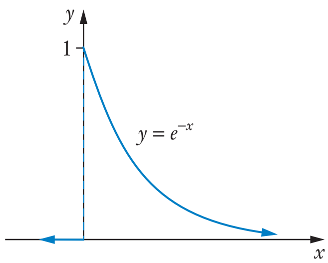


17

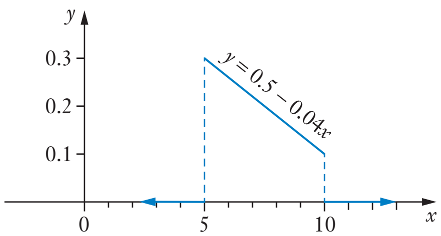


ISBN 9780170395205

18



19



20 The continuous random variable X has the cumulative distribution function

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 5 \\ 0.1(x-5) & \text{for} \quad 5 \le x \le 15 \\ 1 & \text{for} \quad x > 15. \end{cases}$$

Determine: **a** $P(X \le 12)$ **b** $P(X \le 8)$

c $P(8 \le X \le 12)$ **d** $P(X > 8)$

21 Let us suppose that in World War I the number of wartime flying hours a pilot would total before being shot down is a continuous random variable X with cumulative distribution function:

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - e^{-\frac{x}{15}} & \text{for} \quad x \ge 0. \end{cases}$$

Determine, correct to four decimal places, the probability that a World War I pilot:

a is shot down before totalling 20 wartime flying hours,

b totals at least 20 wartime flying hours before being shot down,

c is shot down before reaching a total of 5 wartime flying hours.

d If a pilot manages to total 15 wartime flying hours without being shot down, what is the probability that this pilot makes it to at least 20 hours?

e (Hint: Binomial distribution.)
If we consider 5 World War I pilots, what is the probability that at least 3 of them each totalled at least 20 wartime flying hours before being shot down?

Alamy Stock Photo/Antony Nettle

ISBN 9780170395205

Miscellaneous exercise three

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

1 Solve $3^x - 1 = 5$ giving the *exact* answer using base ten logarithms.

2 A continuous random variable X is uniformly distributed on the interval

$$-2 \le X \le 3.$$

Determine **a** $P(X \ge 0)$ **b** $P(1 \le X \le 2)$ **c** $P(X \le 2 \mid X \ge 1)$

3 If $\log_c 2 = p$ and $\log_c 10 = q$, express each of the following in terms of p or q or both p and q:

a $\log_c 5$ **b** $\log_c 40$ **c** $\log_c 200$

d $\log_c (8c)$ **e** $\log_2 10$ **f** $\log 2$

4 Clearly showing your use of natural logarithms, solve each of the following equations giving your answers as exact values.

a $e^x + e^{x+1} = 17$ **b** $e^{2x+1} = 50^{x-7}$

5 (Revision of the binomial distribution.)

If a normal fair six-sided die is rolled five times what is the probability of obtaining a six on

a exactly three of the rolls?

b all of the first three rolls?

c more than three of the rolls?

d Given that a six occurs more than once in the five rolls, what is the probability that a six occurs more than twice?

Differentiate each of the following with respect to x. For some it may be advisable to use the laws of logarithms *before* differentiating.

6 $y = \ln 5x$ **7** $y = 3x + \ln 3x$ **8** $y = 2\ln x$

9 $y = 2\ln(x^3)$ **10** $y = \ln(2\sqrt{x})$ **11** $y = \ln\left(\frac{2}{x}\right)$

12 Through geological surveys and test drilling, a company discovers a new oil field off the coast of Australia. The experts predict that profitable extraction of the oil can be carried out and that in any one year this extraction will reduce the quantity of oil remaining in the field by 5% of what it was at the beginning of that year. Extraction will become unprofitable when just 20% of the original quantity remains.

For how many years can the company expect the field to remain profitable?

ISBN 9780170395205

13 Find the exact coordinates of the point(s) on the following curves where the gradient is as stated.

a $y = x + \ln 2x$ gradient 1.5.

b $y = \ln[x(x+3)]$ gradient 0.5.

14 The total cost, $\$C$, for producing x units of a certain product is given by:

$$C \approx 600 + 200\ln(1+x).$$

Find **a** an expression for $\frac{dC}{dx}$, the rate of change of C with respect to x,

b the average cost per unit when $\frac{dC}{dx} = 2$.

15 Find the equation of the tangent to $y = \ln(2\sin x)$ at the point $\left(\frac{\pi}{6}, 0\right)$.

16 Let us suppose that for any randomly chosen undecayed atom of a radioactive element, the atom will decay X hours later where X has the cumulative distribution function:

$$\mathrm{P}(X \le x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\frac{x}{8}} & \text{for } x \ge 0. \end{cases}$$

Determine the probability that if we randomly choose an undecayed atom of this radioactive element the atom will decay

a within the next 8 hours,

b within the next twenty-four hours,

c more than twenty-four hours from now.

17 Use calculus to determine the exact coordinates of any stationary points on the graph of $y = 2x^2 - \log_e x$, $(x > 0)$, and use the 2nd derivative test and/or the sign test to determine whether they are maximum, minimum or inflection points.

18 A random variable X has probability density function:

$$f(x) = \begin{cases} ae^{-bx} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Prove that $a = b$.

If $a = 0.25$, find the mean and variance of X, using your calculator to perform appropriate integrations.

ISBN 9780170395205