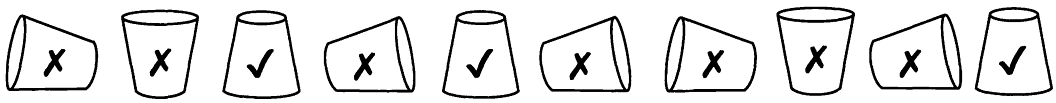
Chapter 8.

Discrete random variables.

Situation One

Suppose we consider it a "success" if, when we drop an empty plastic cup onto a surface, it lands "base up". If you dropped ten such plastic cups onto a surface, how many "successes" would you expect to achieve? Investigate.



Situation Two

Kai and Tenielle are playing a board game. In order to start each of their pieces out from the starting point they must roll a six on a normal, fair, six-sided die. They begin to wonder how many rolls of the die it takes, on average, to get a six. Clearly a six could be achieved in one roll, but it could take a lot more than just one roll.

Ten such results for "rolling a die until a six results" are shown below:

2	2	1	3	6							5 rolls required.
3	2	2	6								4 rolls required.
3	4	1	6								4 rolls required.
5	4	5	2	6							5 rolls required.
5	1	4	2	1	3	6					7 rolls required.
5	3	5	2	3	2	4	4	2	3	6	11 rolls required.
5	1	3	5	4	4	2	4	6			9 rolls required.
1	2	1	4	2	2	2	6				8 rolls required.
6											1 roll required.
5	4	5	3	6							5 rolls required.

Tabulating these results:

Number of rolls required	1	2	3	4	5	6	7	8	9	10	11
Frequency	1	0	0	2	3	0	1	1	1	0	1

We could determine the mean and the standard deviation for this data:

$$\frac{1\times1 + 2\times4 + 3\times5 + 1\times7 + 1\times8 + 1\times9 + 1\times11}{10}$$

Mean = 5·9

$$\sqrt{\frac{1\times(1-5\cdot9)^2 + 2\times(4-5\cdot9)^2 + 3\times(5-5\cdot9)^2 + 1\times(7-5\cdot9)^2 + 1\times(8-5\cdot9)^2 + 1\times(9-5\cdot9)^2 + 1\times(11-5\cdot9)^2}{10}}$$

Standard deviation = 2·737 (correct to 3 decimal places).

Carry out this activity yourself for twenty events of "rolling a die until a six results". Produce a table of results for your twenty occasions like the table shown above and calculate the mean and standard deviation of your results.

Discrete random variables.

Situation One on the previous page asked:

If you dropped ten plastic cups onto a surface, how many "successes" would you expect to achieve? (With a success as defined on the previous page.)

The situation really requires us to carry out the dropping of the ten plastic cups a number of times, to gain some idea of the relative frequencies with which the various possibilities, from zero successes to 10 successes, might occur.

If we use X to represent the number of "successes" obtained in such an experiment, then X can take the values 0, 1, 2, 3, ... 10.

In Situation Two Kai and Tenielle wondered

How many rolls of a die it takes, on average, to get a six.

If we use X to represent the number of rolls it takes to get a six, then X can take the values 1, 2, 3, 4, ...

In each of the above cases the X referred to can take a number of possible integer values, with the value taken dependent on the outcome of a random process. We call X a **discrete random variable**. The word **discrete** means "separate" or "individually distinct" which is the case here because X can only take the distinct values 0, 1, 2, 3, ... 10 in situation one and 1, 2, 3, ... etc., in situation Two. (Contrast this with a situation involving measuring the time a randomly chosen able bodied person takes to run 100 metres. The variable, time in this case, could take any value, between realistic limits. In such a case the random variable would be **continuous**.)

In some cases the frequencies with which the possible values of the random variable may occur are determined empirically, i.e. by actually carrying out the experiment and observing the results, as the previous situations required of you. We can then use the results to estimate probabilities associated with each value of the random variable, or perhaps to estimate the mean value (this would be a **point estimate** because the estimate is a single value, rather than a range of values in which we would expect the mean to lie).

Sometimes the symmetry of the situation, for example the "50-50" nature of a coin flip, allows the theoretical probability of each value occurring to be determined.

For example, for situation two on the previous page, theoretical probabilities could be determined as follows:

$$\text{P(a 6 in 1 roll)} = \frac{1}{6} = \frac{1}{6}$$

$$\text{P(first 6 on 2nd roll)} = \frac{5}{6} \times \frac{1}{6} = \frac{5}{36}$$

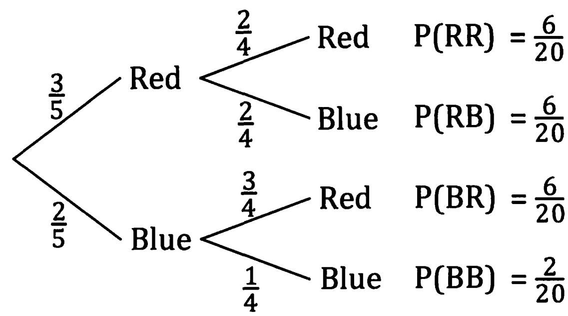
$$\text{P(first 6 on 3rd roll)} = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \frac{25}{216}$$

$$\text{P(first 6 on 4th roll)} = \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \frac{125}{1296}$$

Note:
- The above probabilities form a *geometric progression* with a first term of 1/6 and a common ratio of 5/6.
  What is the *sum to infinity* of such a geometric sequence?

Consider the situation of a bag containing 5 discs, 3 red and 2 blue, indistinguishable except for their colour. Suppose that two discs are randomly selected from the bag, one after the other, the first disc not being returned to the bag before the second is chosen.

The tree diagram for this situation is shown on the right.



If we are interested in the number of red discs this process is likely to select the following table of probabilities would be useful:

Number of red discs	0	1	2
Probability	0·1	0·6	0·3

If we use X to represent the number of reds selected, then X can only take distinct values, in this case the integer values 0, 1 or 2. The value X takes depends on a random process. Thus X is a **discrete random variable** and the table gives the probability associated with each value X can take.

We write $P(X = 0) = 0{\cdot}1$,
$P(X = 1) = 0{\cdot}6$,
$P(X = 2) = 0{\cdot}3$.

☞ The possible values of a random variable must be numerical. (This requirement allows us to consider such things a mean value and standard deviation.)

☞ Some examples of discrete random variables are given below:

Activity.	A possible discrete random variable.	Values variable can take.
• Rolling a normal die.	Number on uppermost face.	1, 2, 3, 4, 5, 6.
• Rolling two normal dice.	Sum of 2 uppermost numbers.	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.
• Flipping a coin n times.	Number of tails obtained.	0, 1, 2, 3, 4, 5, ... , n.
• Cars passing a checkpoint.	Number of cars passing in 5 minutes.	0, 1, 2, 3, 4, 5, 6, 7, ...
• Having 5 children.	Number of boys.	0, 1, 2, 3, 4, 5.
• Repeatedly flipping coin.	Number of flips until a tail is obtained.	1, 2, 3, 4, 5, ...
• Temperature investigation.	Number of days in January with temp ≥ 35°C.	0, 1, 2, 3, ... , 31.
• Subtracting integers.	$a - b$ for $a \in \{1, 2, 3, 4\}$ and $b \in \{1, 2, 3\}$.	−2, −1, 0, 1, 2, 3.
• Attempting 10 questions in a multiple choice test.	Number of questions correct.	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Example 1

Suppose that a fair coin is flipped three times. If X is the discrete random variable "Number of heads obtained" copy and complete the following table.

x	0	1	2	3
$P(X = x)$				

Hence determine (a) $P(X < 1)$,
(b) $P(X \geq 2)$.

Considering the eight equiprobable events:

TTT	HTT THT TTH	THH HTH HHT	HHH
$X = 0$	$X = 1$	$X = 2$	$X = 3$

$P(X = 0) = \frac{1}{8}$ $P(X = 1) = \frac{3}{8}$ $P(X = 2) = \frac{3}{8}$ $P(X = 3) = \frac{1}{8}$

The table can then be completed:

x	0	1	2	3
$P(X = x)$	0·125	0·375	0·375	0·125

(a) $P(X < 1) = 0{\cdot}125$

(b) $P(X \geq 2) = 0{\cdot}375 + 0{\cdot}125$
$= 0{\cdot}5$

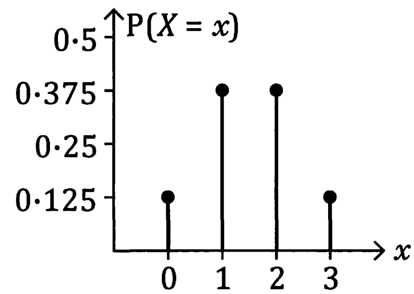
Note

- The table of probabilities completed in the previous example shows how the total probability of 1 is distributed amongst the possible values the random variable X can take. The table gives the **probability distribution** for the random variable X.
- The possible values the random variable can take must together cover all eventualities without overlap. We say they must be **exhaustive** and **mutually exclusive**.
- The sum of the probabilities in a probability distribution must be 1.
- From our understanding of probability it also follows that for each value of x, $0 \leq P(X = x) \leq 1$.
- If the various possible values that the discrete random variable X can take all have the same probability of occurring then we say that the discrete random variable is **uniform**. For the activity of rolling a normal die, if X is the number shown on the uppermost face then X is a **uniform discrete random variable**.
  For a uniform discrete random variable with n possible values, the probability of each value occurring is $\frac{1}{n}$.

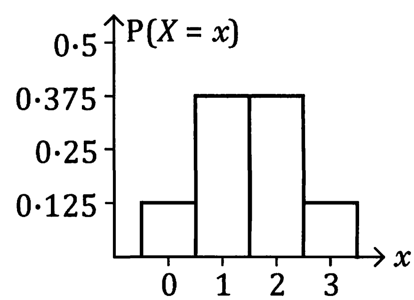
- For each value the discrete random variable, X, can take, the table assigns the corresponding probability of X taking that value. In mathematics we call a rule or relationship that assigns to each element of one set an element from a second set, a **function**. In the previous example the pairs of values in the table show the **probability function** for the random variable X. We frequently use the notation $f(x)$ to represent a function so we will sometimes use $f(x)$ for $P(X = x)$.
- The table below shows the **cumulative probabilities** for example 1. Check that you agree with each of the probabilities.

x	0	1	2	3
$P(X \le x)$	0·125	0·5	0·875	1

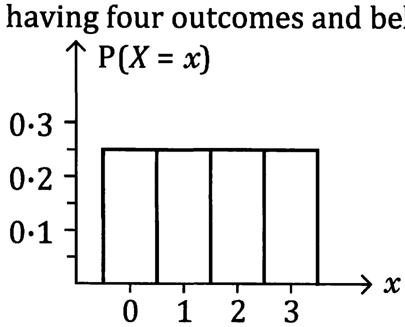
  Note that in a cumulative probability table each probability must be at least as big as the one before it and the final probability must be 1.
- The probability function may be presented as a table showing the pairs of values (x, $P(X = x)$), as in the previous example, and sometimes it may be possible to express the function as a rule, as in the next example.
- Probability functions can also be presented graphically.
  One way this may be done is to use vertical lines, as shown on the right.
  This type of display, with its separated vertical lines, emphasises that we are dealing with a random variable that takes separate values, i.e. a *discrete* random variable.

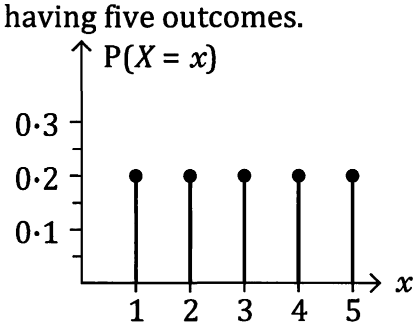


  An alternative form of display that is commonly used shows the distribution as columns of equal width, as shown in the second diagram on the right.
  This second form is sometimes referred to as the *probability histogram*.



The diagrams below are for **uniform random variables** with the one below left having four outcomes and below right having five outcomes.





Example 2

The probability function for a discrete random variable X is given by:

$$P(X = x) = \begin{cases} k(6 - x) & \text{for } x = 1, 2, 3, 4, 5. \\ 0 & \text{for all other values of } x. \end{cases}$$

Copy and complete the following probability distribution for X, giving the probabilities as numbers (i.e. k should be evaluated).

x	1	2	3	4	5
$P(X = x)$					

Determine (a) $P(X = \text{even})$ (b) $P(X > 3)$ (c) $P(X = 4 \mid X > 3)$

$$P(X = 1) = k(6 - 1) = 5k \qquad P(X = 2) = k(6 - 2) = 4k$$

$$P(X = 3) = k(6 - 3) = 3k \qquad P(X = 4) = k(6 - 4) = 2k$$

$$P(X = 5) = k(6 - 5) = k$$

But these probabilities must add up to 1. Thus $15k = 1$

$$\therefore \quad k = \frac{1}{15}$$

The table can then be completed:

x	1	2	3	4	5
$P(X = x)$	$\frac{1}{3}$	$\frac{4}{15}$	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{1}{15}$

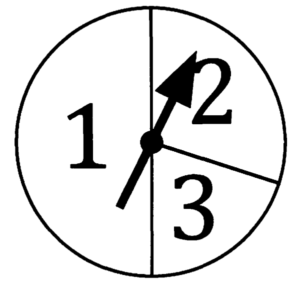
(a) $$P(X = \text{even}) = \frac{4}{15} + \frac{2}{15} = \frac{2}{5}$$

(b) $$P(X > 3) = \frac{2}{15} + \frac{1}{15} = \frac{1}{5}$$

(c) $$P(X = 4 \mid X > 3) = \frac{2}{15} \div \frac{1}{5} = \frac{2}{3}$$

Example 3

A spinner shows the numbers 1, 2, 3. For each spin of this spinner the probability associated with each outcome, 1, 2 or 3, is as shown in the table.



x	1	2	3
$P(X=x)$	0·5	0·3	k

(a) Determine the value of k.

The spinner is spun twice. Determine the probability of getting

(b) a 2 and then a 3,

(c) a 2 and a 3 in any order,

(d) the same number twice,

(e) a total of 4 when the two numbers obtained are added together,

(f) a 2 on the second spin given that the two spins give a total of 4.

(a) The probabilities must sum to 1.

$$\therefore \quad 0{\cdot}5 + 0{\cdot}3 + k = 1$$

$$\text{Thus} \quad k = 0{\cdot}2$$

(b)

$$\begin{aligned} P(2 \text{ then } 3) &= P(2 \text{ on 1st spin}) \times P(3 \text{ on 2nd spin} \mid 2 \text{ on 1st}) \\ &= 0{\cdot}3 \times 0{\cdot}2 \\ &= 0{\cdot}06 \end{aligned}$$

(c)

$$\begin{aligned} P(2 \text{ and } 3 \text{ in any order}) &= P(2 \text{ then } 3) + P(3 \text{ then } 2) \\ &= 0{\cdot}3 \times 0{\cdot}2 + 0{\cdot}2 \times 0{\cdot}3 \\ &= 0{\cdot}12 \end{aligned}$$

(d)

$$\begin{aligned} P(\text{same number twice}) &= P(1 \text{ then } 1) + P(2 \text{ then } 2) + P(3 \text{ then } 3) \\ &= 0{\cdot}5 \times 0{\cdot}5 + 0{\cdot}3 \times 0{\cdot}3 + 0{\cdot}2 \times 0{\cdot}2 \\ &= 0{\cdot}38 \end{aligned}$$

(e)

$$\begin{aligned} P(\text{total of } 4) &= P(1 \text{ then } 3) + P(3 \text{ then } 1) + P(2 \text{ then } 2) \\ &= 0{\cdot}5 \times 0{\cdot}2 + 0{\cdot}2 \times 0{\cdot}5 + 0{\cdot}3 \times 0{\cdot}3 \\ &= 0{\cdot}29 \end{aligned}$$

(f)

$$\begin{aligned} P(2 \text{ second} \mid \text{total of } 4) &= \frac{P(2 \text{ then } 2)}{P(1 \text{ then } 3) + P(3 \text{ then } 1) + P(2 \text{ then } 2)} \\ &= \frac{0{\cdot}09}{0{\cdot}29} \\ &= \frac{9}{29} \end{aligned}$$

Example 4

A batch of 100 components include 5 that are faulty. Four components are randomly selected from the batch without replacement. If X is the number of faulty components in the selection determine the probability distribution for X.

Using ✓ for not faulty and ✗ for faulty:

P(X = 0) = P(✓✓✓✓)

$= \frac{95}{100} \times \frac{94}{99} \times \frac{93}{98} \times \frac{92}{97}$

$\approx$ 0·811 875

P(X = 1) = P(✓✓✓✗) + P(✓✓✗✓) + P(✓✗✓✓) + P(✗✓✓✓)

$= \frac{95}{100} \times \frac{94}{99} \times \frac{93}{98} \times \frac{5}{97} \times 4$

$\approx$ 0·176 495

P(X = 2) = P(✓✓✗✗) + P(✓✗✓✗) + P(✓✗✗✓) + P(✗✓✓✗) + P(✗✓✗✓) + P(✗✗✓✓)

$= \frac{95}{100} \times \frac{94}{99} \times \frac{5}{98} \times \frac{4}{97} \times 6$

$\approx$ 0·011 387

P(X = 3) = P(✓✗✗✗) + P(✗✓✗✗) + P(✗✗✓✗) + P(✗✗✗✓)

$= \frac{95}{100} \times \frac{5}{99} \times \frac{4}{98} \times \frac{3}{97} \times 4$

$\approx$ 0·000 242

P(X = 4) = P(✗✗✗✗)

$= \frac{5}{100} \times \frac{4}{99} \times \frac{3}{98} \times \frac{2}{97}$

$\approx$ 0·000 001

These probabilities are shown tabulated below:

x	0	1	2	3	4
P(X = x)	0·811 875	0·176 495	0·011 387	0·000 242	0·000 001

Alternatively these probabilities can be determined using the notation ^nC_r to represent the number of combinations of r objects chosen from n different objects, as shown on the next page.

As the Preliminary work reminded us:

^nC_r is also written $\binom{n}{r}$, it equals $\frac{n!}{(n-r)!\,r!}$ and can be thought of as "from n *choose* r".

Number of samples of 4 $= {}^{100}C_4$

	Faulty	Okay
From	5	95
Choose	0	4

Number with none faulty $= {}^5C_0 \times {}^{95}C_4$

Thus $P(X=0) = \dfrac{{}^5C_0 \times {}^{95}C_4}{{}^{100}C_4}$

$\approx 0{\cdot}811\,875$

	Faulty	Okay
From	5	95
Choose	1	3

Number with one faulty $= {}^5C_1 \times {}^{95}C_3$

Thus $P(X=1) = \dfrac{{}^5C_1 \times {}^{95}C_3}{{}^{100}C_4}$

$\approx 0{\cdot}176\,495$

	Faulty	Okay
From	5	95
Choose	2	2

Number with two faulty $= {}^5C_2 \times {}^{95}C_2$

Thus $P(X=2) = \dfrac{{}^5C_2 \times {}^{95}C_2}{{}^{100}C_4}$

$\approx 0{\cdot}011\,387$

	Faulty	Okay
From	5	95
Choose	3	1

Number with three faulty $= {}^5C_3 \times {}^{95}C_1$

Thus $P(X=3) = \dfrac{{}^5C_3 \times {}^{95}C_1}{{}^{100}C_4}$

$\approx 0{\cdot}000\,242$

	Faulty	Okay
From	5	95
Choose	4	0

Number with four faulty $= {}^5C_4 \times {}^{95}C_0$

Thus $P(X=4) = \dfrac{{}^5C_4 \times {}^{95}C_0}{{}^{100}C_4}$

$\approx 0{\cdot}000\,001$

Exercise 8A

1. For each of the following state whether the variable is a discrete variable or a continuous variable.
   (a) The heights of students in a year eight class.
   (b) The number of heads obtained in ten flips of a coin.
   (c) The weight of breakfast cereal in packets that each state "contains approximately 500 grammes".
   (d) The number of coins in a piggy bank.
   (e) The number of students in a school.
   (f) The individual weights of 50 dogs seen at a vet's surgery one week.
   (g) The time taken to complete a task.

For each of the following state whether the table of values could represent a probability function for a discrete variable with possible values 0, 1, 2, 3, 4, 5.

2.

x	0	1	2	3	4	5
$f(x)$	0·1	0·2	0·25	0·25	0·1	0·3

3.

x	0	1	2	3	4	5
$f(x)$	0·2	0·1	0·1	0·2	0·1	0·1

4.

x	0	1	2	3	4	5
$f(x)$	−0·2	0·1	0·3	0·5	0·2	0·1

5.

x	0	1	2	3	4	5
$f(x)$	0·2	0·1	0·4	0·1	0·1	0·1

Each of the following tables show probability distributions for the random variable X, with X able to take the values 0, 1, 2, 3, 4. Determine k.

6.

x	0	1	2	3	4
$P(X = x)$	0·1	0·1	0·2	0·2	k

7.

x	0	1	2	3	4
$P(X = x)$	0·25	k	0·25	0·1	0·35

8.

x	0	1	2	3	4
$P(X = x)$	k	2k	2k	3k	2k

9.

x	0	1	2	3	4
$P(X = x)$	−0·25k	k + 0·9	k + 1	−0·5k	k + 0·9

10. Suppose that a fair coin is flipped twice. If X is the discrete random variable "Number of tails obtained", construct a table showing the probability distribution for X.

11. The probability distribution for the random variable X is as shown below:

x	0	1	2	3	4	5
$P(X = x)$	0·2	0·4	0·1	0·1	0·1	0·1

Determine (a) $P(X = 0)$ (b) $P(X \geq 1)$ (c) $P(2 < X \leq 4)$

(d) $P(X = 1 | X \geq 1)$ (e) $P(X > 4 | X \geq 2)$ (f) $P(X \leq 4 | X \geq 2)$

12. The probability distribution for the random variable X is as shown below:

x	0	1	2	3	4	5
$P(X = x)$	0·1	0·2	0·3	0·2	0·1	0·1

Determine (a) $P(X > 2)$ (b) $P(X \geq 3)$ (c) $P(1 < X < 4)$
(d) $P(X = 3 | X > 2)$ (e) $P(X = 5 | X \geq 3)$ (f) $P(X < 4 | X \geq 3)$

13. The discrete random variable X has the following probability distribution:

x	0	1	2	3	4	5	6	7	8	9	10
$P(X = x)$	0·005	0·01	0·04	0·12	0·2	0·25	0·2	0·12	0·04	0·01	0·005

Copy and complete the following table showing cumulative probabilities.

x	0	1	2	3	4	5	6	7	8	9	10
$P(X \leq x)$											

14. The table below shows the cumulative probabilities for the random variable X.

x	0	1	2	3	4	5
$P(X \leq x)$	0·04	0·2	0·5	0·8	0·96	1

Given that the possible values X can take are 0, 1, 2, 3, 4, 5, copy and complete the following table.

x	0	1	2	3	4	5
$P(X = x)$						

15. Suppose that for each flip of a biased coin the probability of getting a head is 0·4. If this coin is flipped twice, and X is the number of tails obtained, copy and complete the following probability distribution table.

x			
$P(X = x)$			

16. Suppose that for each flip of a biased coin the probability of getting a head is $\frac{1}{3}$. If this coin is flipped three times, and X is the number of heads obtained, copy and complete the following probability distribution.

x				
$P(X = x)$				

17. A bag contains 5 marbles, 3 red and 2 blue. Suppose that three marbles are randomly selected from the bag, one after the other, each selected marble not being returned to the bag after selection.
Create a probability distribution table for the random variable X, the number of reds this random selection process produces.

18. The probability function for a discrete random variable X is given by

$$P(X = x) = \begin{cases} kx & \text{for} \quad x = 1, 2, 3, 4, 5. \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Copy and complete the following probability distribution for X, giving the probabilities as numbers (i.e. k should be evaluated).

x	1	2	3	4	5
$P(X = x)$					

Determine (a) $P(X = \text{even})$ (b) $P(X < 2)$ (c) $P(X > 2)$

19. The probability function for a discrete random variable X is given by

$$P(X = x) = \begin{cases} k(5 - x) & \text{for} \quad x = 1, 2, 3, 4. \\ 0 & \text{for} \quad \text{all other values of } x. \end{cases}$$

Copy and complete the following probability distribution for X, giving the probabilities as numbers (i.e. k should be evaluated).

x	1	2	3	4
$P(X = x)$				

Determine (a) $P(X = \text{even})$ (b) $P(X \leq 2)$ (c) $P(X \geq 2)$

20. A spinner shows the numbers 1, 2, 3, 4. For each spin of this spinner the probability associated with each outcome, 1, 2, 3 or 4, is as given in the table below.

x	1	2	3	4
$P(X = x)$	0·2	0·4	0·1	k

(a) Determine the value of k.

The spinner is spun twice. Determine the probability of getting

(b) a 2 and then a 4,
(c) a 2 and a 4 in any order,
(d) a total of 6 when the two numbers obtained are added together,
(e) a 4 on the second spin given that the two spins gave a total of 6.

The spinner is spun three times. Determine the probability of getting

(f) a 4 and then a 3 and then a 2,
(g) a 4, 3 and 2 in any order,
(h) a total of ten when the three numbers obtained are added together,
(i) the same number three times.

21. A batch of 50 components include 5 that are faulty. Four components are randomly selected from the batch. If X is the number of faulty components in the selection, determine the probability distribution for X, giving probabilities correct to five decimal places.

Mean or expected value of a discrete random variable.

Consider the probability distribution for rolling a normal, fair, six-sided die.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

If we were to roll such a die 6000 times we would expect to obtain roughly 1000 of each of the six possible numbers. Hence we would expect our mean to be approximately:

$$\frac{1000 \times 1 + 1000 \times 2 + 1000 \times 3 + 1000 \times 4 + 1000 \times 5 + 1000 \times 6}{6000} = 3{\cdot}5$$

Writing the above calculation of the mean as follows

$$\frac{1000 \times 1}{6000} + \frac{1000 \times 2}{6000} + \frac{1000 \times 3}{6000} + \frac{1000 \times 4}{6000} + \frac{1000 \times 5}{6000} + \frac{1000 \times 6}{6000}$$

this is the same as

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}.$$

Thus we could obtain the mean value by summing the products:

(value of the random variable) × (probability of that value occurring).

When working with random variables the mean value is sometimes referred to as the **expected value**. For the random variable X, the mean or expected value is sometimes written as $E(X)$.

Note **Note**

Do not be misled by the use of the word "expected". Clearly 3·5 is not the outcome we "expect" from one roll of a normal die. (Just as "two and a bit children" is not the number of children we expect to find in a randomly chosen family.) Hence the expected value is not the value we expect to get with one roll of the die but is instead the number we expect our long term average to be close to.

Note **Note**

For the following probability distribution for the random variable X,

x	1	2	3	4	5
$P(X = x)$	0·1	0·2	0·2	0·4	0·1

the mean or expected value $= 1 \times 0{\cdot}1 + 2 \times 0{\cdot}2 + 3 \times 0{\cdot}2 + 4 \times 0{\cdot}4 + 5 \times 0{\cdot}1$

$= 3{\cdot}2$

If the discrete random variable, X, has possible values x_i, with $P(X = x_i) = p_i$ then

$$E(X) = \sum(x_i \, p_i)$$

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

Example 5

A game involves a player paying money to roll two dice.

If the total of the two numbers on the uppermost faces is

- less than 6, then the player receives \$5,
- more than 9, then the player receives \$8,
- any other score, then the player receives nothing.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

(a) What is the amount the player could expect to receive, on average, per game (i.e. what is the mean or expected payout per game)?

(b) If the cost for each roll of the two dice is \$3, is a player likely to win, lose or break even in the long term?

(a) If X is the sum of the two uppermost numbers the probability distribution for X is as follows:

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
	Player receives \$5.				Player receives nothing.				Player receives \$8.		

Hence if Y is the number of dollars the player receives the probability distribution for Y is as follows:

y (dollars received)	0	5	8
$P(Y = y)$	$\frac{20}{36}\left(=\frac{5}{9}\right)$	$\frac{10}{36}\left(=\frac{5}{18}\right)$	$\frac{6}{36}\left(=\frac{1}{6}\right)$

Expected payout per game $= \$0 \times \frac{5}{9} + \$5 \times \frac{5}{18} + \$8 \times \frac{1}{6}$

$= \$2{\cdot}72$ (to the nearest cent)

(b) If each game costs \$3 to play and the average pay out is \$2·72 per game the player is likely to lose in the long term. (For example playing 100 games would cost the player \$300 for an expected return of \$272 - a loss of \$28.)

Note • The game would be regarded as being "fair" if the cost to play equalled the expected pay out per game.

The standard deviation of a discrete random variable.

Remember that to determine the standard deviation of a set of scores we find the deviation of each score from the mean, find the average of the squares of these deviations (this gives us the variance of the scores) and then square root our answer to obtain the standard deviation.

Thus, as the Preliminary Work reminded us, for the following 8 scores (for which the mean is 18):

	12	15	16	16	18	20	22	25
Deviation from mean:	−6	−3	−2	−2	0	+2	+4	+7

$$\text{Variance of scores} = \frac{(-6)^2 + (-3)^2 + (-2)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2 + (7)^2}{8}$$

$$= 15{\cdot}25$$

$$\text{Standard deviation} = \sqrt{15{\cdot}25} \quad \text{i.e. } 3{\cdot}91 \text{ (correct to two decimal places)}$$

Similarly, if the discrete random variable X has possible values x_i with $P(X = x_i) = p_i$ and a mean, or expected value $E(X)$, of μ then

The variance, sometimes written $Var(X)$, $= \sum[p_i (x_i - \mu)^2]$ and the standard deviation is the square root of the variance.

For the dice rolling situation:

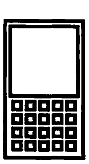
x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

As determined earlier, $E(X) = 3{\cdot}5$. Hence the variance of the distribution is given by

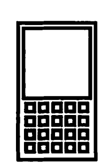
$$\frac{1}{6}\times(1-3{\cdot}5)^2 + \frac{1}{6}\times(2-3{\cdot}5)^2 + \frac{1}{6}\times(3-3{\cdot}5)^2 + \frac{1}{6}\times(4-3{\cdot}5)^2 + \frac{1}{6}\times(5-3{\cdot}5)^2 + \frac{1}{6}\times(6-3{\cdot}5)^2$$

$$= \frac{35}{12}$$

The standard deviation will then be $\sqrt{Var(X)}$, i.e. 1·7078, correct to 4 decimal places.
The standard deviation of X is sometimes written $SD(X)$.



Of course, rather than performing these calculations to determine the mean, or expected value, and the standard deviation of a discrete random variable "by hand", we can use the statistical capability of a calculator. Considering the probability as the frequency of occurrence, use the statistics facility of a calculator to output the expected value and the standard deviation for the dice rolling probability distribution given above.



Example 6

The discrete random variable X can take the values 1, 2, 3 and 4 with the probability distribution as given in the table below.

x	1	2	3	4
$P(X = x)$	0·1	0·2	0·4	0·3

(a) Find $E(X)$, the expected value of X, and $Var(X)$, the variance of X.

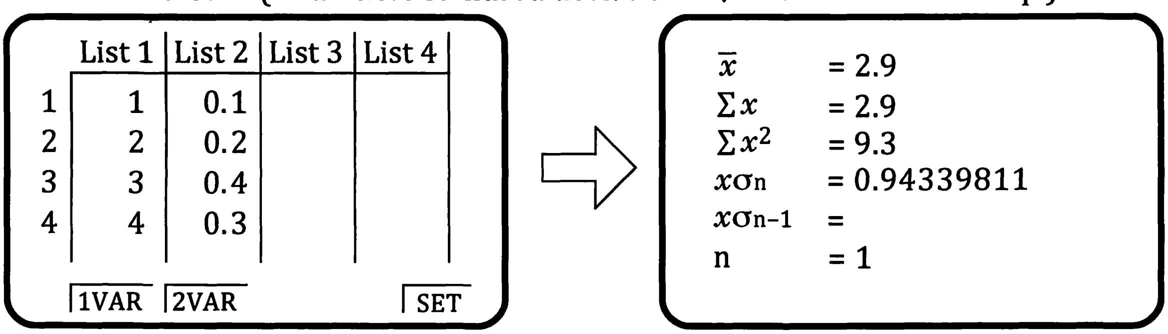
(b) If $Y = 3X$ find $E(Y)$, the expected value of Y, and $Var(Y)$, the variance of Y.

(b) If $Z = 3X + 4$ find $E(Z)$, the expected value of Z, and $Var(Z)$, the variance of Z.

(a) Either by using the statistical functions of a calculator, or by calculation, as shown below,

$$E(X) = 1 \times 0{\cdot}1 + 2 \times 0{\cdot}2 + 3 \times 0{\cdot}4 + 4 \times 0{\cdot}3$$
$$= 2{\cdot}9$$

$$Var(X) = 0{\cdot}1 \times (1 - 2{\cdot}9)^2 + 0{\cdot}2 \times (2 - 2{\cdot}9)^2 + 0{\cdot}4 \times (3 - 2{\cdot}9)^2 + 0{\cdot}3 \times (4 - 2{\cdot}9)^2$$
$$= 0{\cdot}89 \quad \text{(And hence standard deviation} = \sqrt{0{\cdot}89} = 0{\cdot}9434 \text{ to 4 dp.)}$$



(b) With $Y = 3X$ we now have the distribution:

y	3	6	9	12
$P(Y = y)$	0·1	0·2	0·4	0·3

We have a *change of scale*, as referred to in the Preliminary work.
This will multiply the mean value by 3 and the standard deviation by 3 (and hence the variance by 3^2).

Thus $E(Y) = 3 \times E(X) = 3 \times 2{\cdot}9 = 8{\cdot}7$ and $Var(Y) = 3^2 \times Var(X) = 9 \times 0{\cdot}89 = 8{\cdot}01$

(c) With $Z = 3X + 4$ we have both a change of scale and a change of origin. The change of scale alters both the mean and the standard deviation whereas the change of origin does not alter the standard deviation (because it does not alter the spread).

Thus $E(Z) = 3 \times E(X) + 4 = 3 \times 2{\cdot}9 + 4 = 12{\cdot}7$ and $Var(Z) = 3^2 \times Var(X) = 9 \times 0{\cdot}89 = 8{\cdot}01$

Note: (For interest.) Prior to the ready availability of calculators with statistical capabilities, calculating the variance, and the standard deviation, of a probability distribution could be a tedious process, especially if E(*X*) was not an integer. Subtracting each value from E(*X*), in order to calculate variance, was a chore. In such cases, the alternative formula $\text{Var}(X) = \text{E}(X^2) - [\text{E}(X)]^2$ could be used. The reader might like to check that applying this formula to the distribution for *X* given in the previous example also gives Var(*X*) = 0·89.

Exercise 8B.

The tables in number 1 to 4 show discrete probability distributions for the discrete random variable *X*.
Find k and the mean (or expected) value of *X* in each case.

1.

x	1	2	3	4	5
$P(X = x)$	k	0·35	0·35	0·15	0·05

2.

x	0	5	10	15	20	25
$P(X = x)$	0·1	0·1	0·1	0·1	k	2k

3.

x	1	2	3	4	5	6	7	8
$P(X = x)$	k	k	2k	k	2k	3k	4k	6k

4.

x	1	2	5	8	10	20	25	50	100
$P(X = x)$	0·05	0·15	k	0·25	0·15	0·10	0·05	0·03	0·02

5. The expected value of the discrete probability distribution given below is 2·7. Find the values of p and q and hence determine Var(*X*), the variance of *X*.

x	1	2	3	4	5
$P(X = x)$	0·3	p	0·2	q	0·1

6. The expected value of the discrete probability distribution given below is $\frac{52}{9}$. Find the values of p and q.

x	0	1	2	3	4	5	6	7	8
$P(X = x)$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	p	q

7. If the discrete random variable X has an expected value, $E(X)$, of 10 and a standard deviation of 1·5, find
   (a) $E(X + 5)$,
   (b) the standard deviation of $(X + 5)$,
   (c) $E(3X - 4)$,
   (d) the standard deviation of $(3X - 4)$.

8. The probability distribution of the discrete random variable X is as shown below.

x	10	20	30	40	50
$P(X = x)$	0·3	0·2	0·2	0·2	0·1

(a) Find $E(X)$, the expected value of X, and $Var(X)$, the variance of X.
Hence write the value of each of the following:
(b) $E(X + 3)$ (c) $E(2X)$ (d) $E(2X + 3)$
(e) $Var(X + 3)$ (f) $Var(2X)$ (g) $Var(2X + 3)$

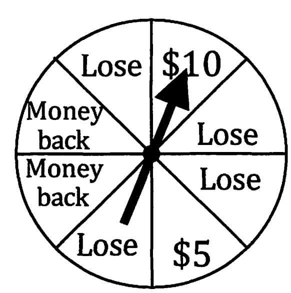
9. A <u>uniform</u> discrete random variable, X, can take the values 1, 2, 3, 4, 5.
Find $E(X)$ and $Var(X)$.

10. Sue and Bob are developing a gambling game. Each "play" of the game involves the rolling of two normal fair six sided dice and the two numbers on the uppermost faces being added together.
Any double, i.e. two 1s, two 2s, two 3s etc., pays \$30.
A total of 7 pays \$15.
Anything other than the above two outcomes pays nothing.
If the discrete random variable, \$$X$, is the amount paid out on a single play, copy and complete the following probability distribution table for X.

x (\$ paid out)	0	15	30
$P(X = x)$			

In the long run, Sue and Bob want to make an average of \$0·50 profit per game played. How much should they charge for each "play"?

11. A game is being devised based on the spinner shown on the right. The spinner features eight equal size sectors and each spin will be such that each sector can be considered to be equally likely to be the one the arrow finally points to.
A player pays an amount to play the game and each play involves one spin of the spinner.
Each sector indicates the amount received should the arrow end up pointing to that sector with "lose" meaning nothing is received and "money back" meaning the cost of the game is paid back.
How much should the organisers charge for each game so that in the long run they should at least break even (or be very close to break even)?



12. A fair eight sided die featuring the numbers 1, 2, 3, 4, 5, 6, 7 and 8 with one number on each face, is rolled.

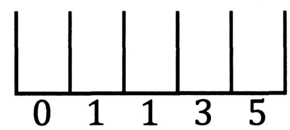
    (a) The discrete random variable X is the number shown on the uppermost face. Find the mean value of X.

    (b) The discrete random variable Y is the square of the number on the uppermost face. Find the mean value of Y.

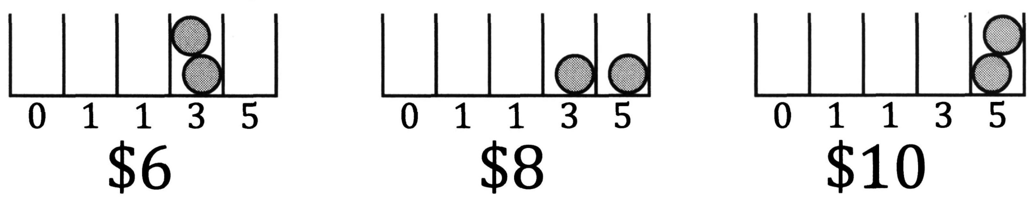
    (c) The discrete random variable Z is $\dfrac{1}{\text{the number on the uppermost face}}$.

    Find the mean value of Z.

13. A game involves two balls being released, one after the other, rolling down a slope and each ending up in one of five numbered slots, as shown on the right. The random release mechanism is such that each slot has an equal chance of receiving each ball.



A player pays \$5 for one release of the two balls. The total of the scores achieved by the two balls is the player's score for the game and the player receives that number of dollars as a prize.



(a) What percentage of games would you expect to result in a prize of more than \$6 being awarded? (i.e. a profit to the player of more than \$1.)
(b) What total score would we expect 40% of the plays to exceed?
(c) Approximately how much should the organisers expect to be "up" after 100 plays of the game?

14. Bill is offered a part time job as a salesman of new cars. He is told that an analysis of the data from other sales people indicates that if X is the number of new car sales he can expect to achieve in a fortnight, when working the proposed number of hours, the probability distribution for X tends to approximately follow the pattern:

x	0	1	2	3	4	5
$P(X = x)$	0·15	0·25	0·35	0·15	0·05	0·05

He is offered two schemes of reward:

① \$500 per fortnight plus \$250 for each new car sold,

or ② \$nil per fortnight plus \$475 per new car sold.

Based on expected earnings which scheme should he prefer? (Justify your answer.)

15. (a) A financial analyst estimates that, for a particular investment scheme, if $\$X$ is the value after three years of an initial investment of \$1000 then X has a probability distribution:

x	500	800	1000	2000	3000
$P(X=x)$	0·2	0·3	0·1	0·3	0·1

Determine the mean, or expected, value of the initial \$1000 at the end of the three years.

(b) An alternative investment offers $\$Y$ as the final value after three years of an initial investment of \$1000 with the same analyst estimating the probability distribution for Y as:

y	800	1000	1200	1500
$P(Y=y)$	0·1	0·2	0·2	0·5

Determine the mean, or expected, value of the initial \$1000 at the end of the three years.

(c) Which of the two schemes would you advise is the better and why?

Miscellaneous Exercise Eight.

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. A medical research team monitoring the spread of a particular notifiable disease in a certain country town surveyed the number of cases that the authorities were notified about over a period of time. The total number of notified cases, t weeks after the survey commenced was N where:

$$N \approx \frac{100\,000}{1 + 499e^{-0\cdot08t}}.$$

Determine (a) the number of recorded cases when the survey commenced,
(b) the number of recorded cases when $t = 5$,
(c) the number of recorded cases when $t = 10$,
(d) what happens as $t \to \infty$ if no action occurs to inhibit the spread of the disease?

2. Differentiate each of the following expression with respect to x.

(a) $\dfrac{6}{x}$ (b) $\dfrac{6}{\sqrt{x}}$

(c) $5x^2 - e^x$ (d) $e^{3x^2} + 1$

(e) e^{3x^2+1} (f) $(2x-3)(2x+1)^5$

(g) $10 \sin x$ (h) $\sin 10x$

3. Determine $\dfrac{d}{dx}\left(\displaystyle\int_0^x (3t^2 - 5)\,dt\right)$.

4. For each of the following state whether the variable X is a uniform discrete random variable and explain your reasoning for each one.
   (a) X is the number of sixes obtained when two normal fair dice are rolled.
   (b) X is the number on the upper most face of a normal fair die when it is rolled.
   (c) X is the height of a randomly chosen male aged between 20 and 25.

5. Find the coordinates of any points where the curve $y = \dfrac{5x^2}{x-1}$
   cuts the line $y = 5x + 3$.

   Find the gradient of the curve at these points.

6. Find the exact gradient of $y = x^2e^{2x}$ at the point $(1, e^2)$.

7. Find the equation of the normal to $y = \dfrac{x^2}{x-2}$ at the point (3, 9).

   (Note: • The *normal* to a curve, at some point A on the curve, is the line through A perpendicular to the tangent to the curve at point A.
   • The gradients of two perpendicular lines have a product of −1.)

8. Evaluate the following definite integrals without the assistance of your calculator.

   (a) $\displaystyle\int_0^2 10x^4\,dx$ (b) $\displaystyle\int_2^4 2\,dx$ (c) $\displaystyle\int_2^3 (2 + 6x)\,dx$

9. Find the area between $y = \sin x$ and the x-axis from $x = 0$ to $x = \dfrac{\pi}{2}$.

10. Let us suppose that a particular freefalling object, under specific conditions of air resistance etc., falls such that its downward velocity, t seconds after release, is given by v m/sec, where $v = 75(1 - e^{-0\cdot13t})$ m/sec.
    (a) An object in free fall does not keep accelerating indefinitely but, due to air resistance, reaches a terminal velocity. What is the terminal velocity of this free falling object?
    (b) Find, in m/s^2 and correct to two decimal places, the acceleration of the object
        (i) 5 seconds after release,
        (ii) 20 seconds after release.

11. Find $f(x)$ given that $f''(x) = 20(3 - x)^3 + 6x - 6$, $f'(1) = -83$ and $f(1) = 28$.

12. By writing x^5 as $\dfrac{x^3 \times x^4}{x^2}$ differentiate $y = x^5$ using the product and quotient rules.

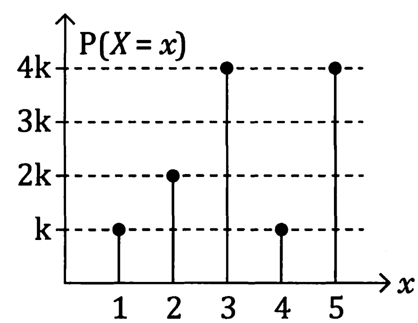
13. Find the exact coordinates of any points on $y = \dfrac{e^x}{2x}$ where the gradient is 0.

14. Find the exact values of x, $-2\pi \leq x \leq 2\pi$, for which the curve $y = e^x \cos x$ has a gradient of zero.

15. Integrate the following with respect to x without the assistance of your calculator.

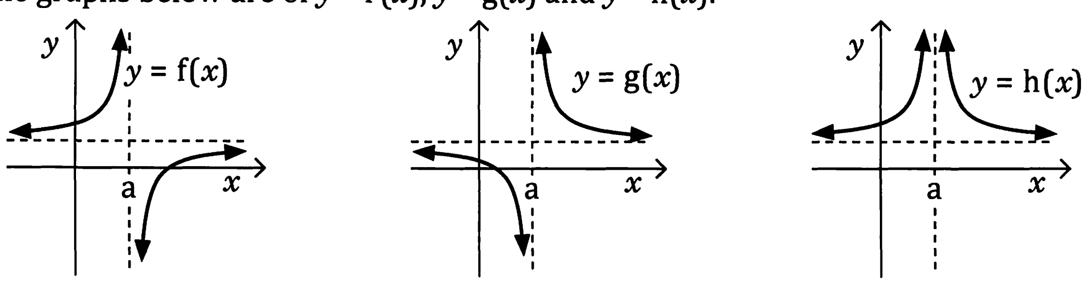
(a) $4x$ (b) $6e^{2x}$ (c) $\dfrac{d}{dx}(x^2 + e^x)$ (d) $\dfrac{d}{dx}(x^2 e^x)$

16. The discrete random variable X can take the values 1, 2, 3, 4, 5. The probability distribution for X is shown graphically on the right.



Determine (a) k,
(b) $P(X = 3)$,
(c) $P(X > 3)$,
(d) $P(X \geq 3)$,
(e) $P(X = 3 | X > 3)$,
(f) $P(X = 3 | X \geq 3)$.
(g) Without the use of your calculator, determine E(X), the expected value of X.
(h) With the assistance of your calculator, determine the standard deviation of X, giving your answer correct to two decimal places.

17. The graphs below are of $y = f(x)$, $y = g(x)$ and $y = h(x)$.

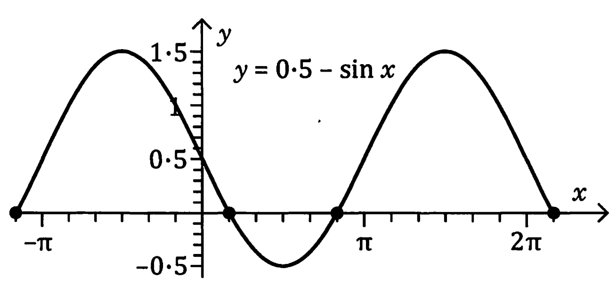


Produce sketches showing the graphs of $y = f'(x)$, $y = g'(x)$ and $y = h'(x)$.

18. Find as exact values

(a) $\displaystyle\int_0^{5\pi/6} (0{\cdot}5 - \sin x)\ dx$

(b) $\displaystyle\left| \int_0^{5\pi/6} (0{\cdot}5 - \sin x)\, dx \right|$



(c) the area between $y = 0{\cdot}5 - \sin x$ and the x-axis from $x = 0$ to $x = 5\pi/6$.