4

PROBABILITY DISTRIBUTIONS I — DISCRETE RANDOM VARIABLES

If we toss a coin twice, the number of heads obtained could be 0, 1 or 2. The probabilities of these occurring are as follows:

$$P(\text{no heads}) = P(TT) = (0.5)(0.5) = 0.25$$

$$P(\text{one head}) = P(HT) + P(TH) = (0.5)(0.5) + (0.5)(0.5) = 0.5$$

$$P(\text{two heads}) = P(HH) = (0.5)(0.5) = 0.25$$

We can show the results in a table, known as a **probability distribution**.

Number of heads	0	1	2
Probability	0.25	0.5	0.25

We now develop some useful notation.

The variable being considered is 'the number of heads obtained in two tosses' and it can be denoted by X. It can only take exact values, 0, 1 and 2 and so is called a **discrete** variable.

The probabilities can be written

$$P(X = 0) = 0.25, \ P(X = 1) = 0.5, \ P(X = 2) = 0.25$$

Sometimes we write $p_0 = 0.25, \ p_1 = 0.5, \ p_2 = 0.25$.

Now if the sum of the probabilities is 1, the variable is said to be **random**.

In this example

$$P(X = 0) + P(X = 1) + P(X = 2) = 0.25 + 0.5 + 0.25 = 1,$$

so X is a discrete random variable.

The probability distribution is often written

x	0	1	2
$P(X = x)$	0.25	0.5	0.25

and the statement 'the sum of the probabilities is 1' is written

$$\sum_{\text{all } x} P(X = x) = 1$$

DISCRETE RANDOM VARIABLE

Let X have the following properties:

(a) it is a discrete variable and can take only values $x_1, x_2, \ldots, x_n$;

(b) the probabilities associated with these values are $p_1, p_2, \ldots, p_n$,

where
$$P(X = x_1) = p_1$$
$$P(X = x_2) = p_2$$
$$\vdots$$
$$P(X = x_n) = p_n.$$

Then X is a discrete random variable if $p_1 + p_2 + \ldots + p_n = 1$.

This can be written

$$\sum p_i = 1, \qquad i = 1,2,\ldots,n$$

or

$$\sum_{\text{all } x} P(X = x) = 1$$

We usually denote a random variable (r.v.) by a capital letter (X, Y, R, etc.) and the particular value it takes by a small letter (x, y, r, etc.).

Example 4.1 Let X be the discrete variable 'the number of fours obtained when two dice are thrown'. Show that X is a random variable, i.e. that the sum of the probabilities is 1. Illustrate the probability distribution on a diagram.

Solution 4.1 When two dice are thrown, the number of fours obtained is 0, 1 or 2. Therefore X can take the values 0, 1 and 2 only.

Then, with obvious notation,

$$P(X=0) = P(\bar{4}\,\bar{4}) = \left(\tfrac{5}{6}\right)\left(\tfrac{5}{6}\right) = \tfrac{25}{36}$$

$$P(X=1) = P(4\,\bar{4}) + P(\bar{4}\,4) = \left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right) + \left(\tfrac{5}{6}\right)\left(\tfrac{1}{6}\right) = \tfrac{10}{36}$$

$$P(X=2) = P(4\,4) = \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) = \tfrac{1}{36}$$

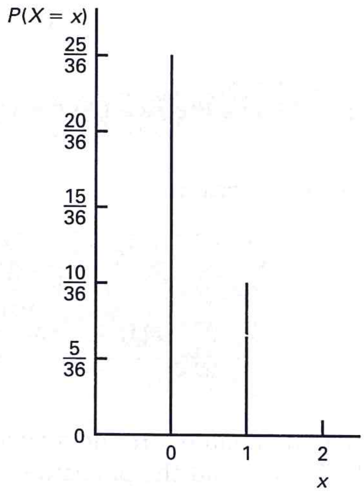
Now
$$\sum_{\text{all } x} P(X=x) = \tfrac{25}{36} + \tfrac{10}{36} + \tfrac{1}{36} = \tfrac{36}{36} = 1$$

Therefore X is a random variable.

The probability distribution is

x	0	1	2
$P(X=x)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

and it can be represented by a vertical line graph:



PROBABILITY DENSITY FUNCTION (p.d.f.)

The function that is responsible for allocating probabilities, $P(X=x)$, is known as the **probability density function** (p.d.f.) of X.

Sometimes it can be expressed as a formula, as in the following example.

Example 4.2 Two tetrahedral dice, each with faces labelled 1, 2, 3 and 4 are thrown and the score noted, where the score is the sum of the two numbers on which the dice land. Find the probability density function (p.d.f.) of X, where X is the random variable 'the score when two dice are thrown'.

Solution 4.2 The score for each possible outcome is shown in the table:

'Score'

Second die \ First die	1	2	3	4
4	5	6	7	8
3	4	5	6	7
2	3	4	5	6
1	2	3	4	5

From the table we can see that X can take the values 2, 3, 4, 5, 6, 7, 8 only.

The probabilities can be found from the table, since each outcome shown is equally likely.

For example, $P(X = 5) = \frac{4}{16}$ since 4 out of the total of 16 outcomes result in a score of 5.

The probability distribution is formed:

x	2	3	4	5	6	7	8
$P(X = x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

This can be written as a formula, giving the p.d.f. of X as

$$P(X = x) = \frac{x-1}{16} \quad \text{for} \quad x = 2, 3, 4, 5$$

$$P(X = x) = \frac{9-x}{16} \quad \text{for} \quad x = 6, 7, 8$$

NOTE: $\sum_{\text{all } x} P(X = x) = \frac{1}{16}(1+2+3+4+3+2+1) = 1,$

confirming that X is a random variable.

Example 4.3 The p.d.f. of a discrete random variable Y is given by $P(Y = y) = cy^2$, for $y = 0, 1, 2, 3, 4$. Given that c is a constant, find the value of c.

Solution 4.3 The probability distribution of Y is

y	0	1	2	3	4
$P(Y = y)$	0	c	$4c$	$9c$	$16c$

Since Y is a random variable, $\sum_{\text{all } y} P(Y = y) = 1.$

So
$$c + 4c + 9c + 16c = 1$$
$$30c = 1$$
$$c = \frac{1}{30}$$

Therefore, since Y is a random variable, $\underline{c = \frac{1}{30}}$.

Example 4.4 The p.d.f. of the discrete r.v. is given by $P(X = x) = a\left(\frac{3}{4}\right)^x$ for $x = 0, 1, 2, 3, \ldots$. Find the value of the constant, a.

Solution 4.4 Since X is a random variable, $\sum_{\text{all } x} P(X = x) = 1$.

Now
$$P(X = 0) = a\left(\frac{3}{4}\right)^0$$
$$P(X = 1) = a\left(\frac{3}{4}\right)^1$$
$$P(X = 2) = a\left(\frac{3}{4}\right)^2$$
$$P(X = 3) = a\left(\frac{3}{4}\right)^3$$

and so on.

So
$$\sum_{\text{all } x} P(X = x) = a + a\left(\frac{3}{4}\right) + a\left(\frac{3}{4}\right)^2 + a\left(\frac{3}{4}\right)^3 + \ldots$$
$$= a\left(1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \left(\frac{3}{4}\right)^3 + \ldots\right)$$
$$= a\left(\frac{1}{1 - \frac{3}{4}}\right) \quad \left(\text{sum of an infinite G.P. with first term 1 and common ratio } \frac{3}{4}\right)$$
$$= a(4)$$

We have
$$4a = 1$$

Therefore
$$\underline{a = \frac{1}{4}}$$

Example 4.5 The discrete random variable W has p.d.f. as shown

w	-3	-2	-1	0	1
$P(W = w)$	0.1	0.25	0.3	0.15	d

Find (**a**) the value of d, (**b**) $P(-3 \leqslant W < 0)$, (**c**) $P(W > -1)$, (**d**) $P(-1 < W < 1)$, (**e**) the mode.

Solution 4.5 (a) Now $\sum_{\text{all } w} P(W = w) = 1$

so
$$0.1 + 0.25 + 0.3 + 0.15 + d = 1$$
$$0.8 + d = 1$$
$$\underline{d = 0.2}$$

(b)
$$P(-3 \leqslant W < 0) = P(W = -3) + P(W = -2) + P(W = -1)$$
$$= 0.1 + 0.25 + 0.3$$
$$= \underline{0.65}$$

(c)
$$P(W > -1) = P(W = 0) + P(W = 1)$$
$$= 0.15 + 0.2$$
$$= \underline{0.35}$$

(d)
$$P(-1 < W < 1) = P(W = 0)$$
$$= \underline{0.15}$$

(e) The value of w that has the highest probability is -1.
Therefore $\underline{\text{the mode} = -1.}$

Exercise 4a

1. The discrete random variable X has p.d.f. as shown

x	1	2	3	4	5
$P(X = x)$	0.2	0.25	0.4	a	0.05

Find (i) the value of a
(ii) $P(1 \leqslant X \leqslant 3)$
(iii) $P(X > 2)$
(iv) $P(2 < X < 5)$
(v) the mode.

Draw a vertical line graph to illustrate the distribution.

2. The probability density function of a discrete random variable X is given by $P(X = x) = kx$ for $x = 12, 13, 14$. Find the value of the constant k.

3. The discrete random variable R has p.d.f. given by $P(R = r) = c(3 - r)$ for $r = 0, 1, 2, 3$. Find the value of the constant c, and draw a vertical line graph to illustrate the distribution.

4. For each of the following random variables write out the probability distributions. Check that the variables are random and for parts (b), (d) and (f) write the formula for the p.d.f.
(a) The number of heads obtained when two fair coins are tossed.
(b) The sum of the scores when two ordinary dice are thrown.
(c) The number of threes obtained when two tetrahedral dice are thrown.
(d) The numerical value of a digit chosen from a set of random number tables.
(e) The number of tails obtained when three fair coins are tossed.
(f) The difference between the numbers when two ordinary dice are thrown.

5. A drawer contains 8 brown socks and 4 blue socks. A sock is taken from the drawer at random, its colour is noted and it is then replaced. This procedure is performed twice more. If X is the r.v. 'the number of brown socks taken', find the probability distribution for X.

6. The r.v. X has p.d.f. $P(X = x) = c\left(\frac{4}{5}\right)^x$ for $x = 0, 1, 2, 3, \ldots$. Find the value of the constant, c.

7. A game consists of throwing tennis balls into a bucket from a given distance. The probability that William will get the tennis ball in the bucket is 0.4. A turn consists of three attempts.
(i) Construct the probability distribution for X, the number of tennis balls that land in the bucket in a turn.
(ii) William wins a prize if, at the end of his turn, there are two or more tennis balls in the bucket. What is the probability that William does not win a prize?

8. A student has a fair coin and two six-sided dice, one of which is white and the other blue. The student tosses the coin and then rolls both dice. Let X be a random variable such that if the coin falls heads, X is the sum of the scores on the two dice, otherwise X is the score on the white die only.
Find the probability function of X in the form of a table of possible values of X and their associated probabilities.
Find $P(3 \leqslant X \leqslant 7)$.
State the assumption you made to enable you to evaluate the probability function.
(AEB 1991)P

EXPECTATION, $E(X)$

Experimental approach

Suppose we throw an unbiased die 120 times and record the results in a **frequency distribution:**

Score, x	1	2	3	4	5	6	
Frequency, f	15	22	23	19	23	18	Total 120

We can calculate the mean score obtained as follows:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{15 + 44 + 69 + 76 + 115 + 108}{120} = 3.558 \quad \text{(3 d.p.)}$$

Theoretical approach

The **probability distribution** for the random variable X, where X is 'the number on the die', is as shown:

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

We can obtain a value for the 'expected' mean by multiplying each score by its corresponding probability and summing.

$$\begin{aligned}\text{Expected mean} &= 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) \\ &= \frac{21}{6} \\ &= 3.5\end{aligned}$$

If we have a statistical experiment:

a practical approach results in a frequency distribution and a mean value,

a theoretical approach results in a probability distribution and an expected value, known as the **expectation**.

The expectation of X (or expected value), written $E(X)$, is given by

$$E(X) = \sum_{\text{all } x} xP(X = x)$$

This can also be written

$$E(X) = \sum x_i p_i \qquad i = 1, 2, \ldots, n$$

We often use the symbol μ, pronounced 'mew', for the expectation, so

$$\mu = E(X)$$

Example 4.6 A random variable X has probability density function (p.d.f.) as shown. Find the expectation, $E(X)$.

x	-2	-1	0	1	2
$P(X = x)$	0.3	0.1	0.15	0.4	0.05

Solution 4.6

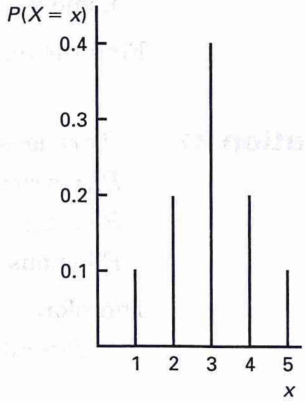
$$\begin{aligned} E(X) &= \sum_{\text{all } x} xP(X = x) \\ &= (-2)(0.3) + (-1)(0.1) + 0(0.15) + 1(0.4) + 2(0.05) \\ &= -0.2 \end{aligned}$$

Therefore $E(X) = -0.2$.

NOTE: an important property which some probability distributions possess is that of symmetry. For example,

(a) Consider the r.v. with probability distribution:

x	1	2	3	4	5
$P(X = x)$	0.1	0.2	0.4	0.2	0.1

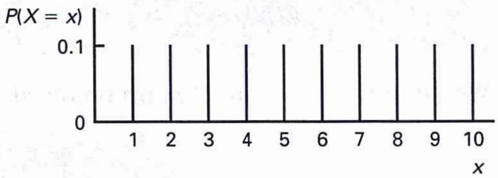


It can be seen from the table or from the vertical line graph that the distribution is symmetrical about the central value $X = 3$, so $E(X) = 3$.

Check: $E(X) = \sum_{\text{all } x} xP(X = x) = 1(0.1) + 2(0.2) + 3(0.4) + 4(0.2) + 5(0.1) = 3$

(**b**) Consider the r.v. with p.d.f. $P(X = x) = 0.1$ for $x = 1,2,\ldots,10$.
The probability distribution for X is

x	1	2	3	4	5	6	7	8	9	10
$P(X = x)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1



The distribution is symmetrical about the central value mid-way between 5 and 6, so $E(X) = 5.5$.

NOTE: the random variable X with p.d.f. $P(X = x) = k$, for all given values of x, where k is a constant, is said to follow a **uniform distribution**.

Example 4.7 A fruit machine consists of three windows, each of which shows pictures of fruits — lemons or oranges or cherries or plums. The probability that a window shows a particular fruit is as follows:

$P(\text{lemons}) = 0.4$, $P(\text{oranges}) = 0.1$, $P(\text{cherries}) = 0.2$, $P(\text{plums}) = 0.3$

The windows operate independently.

Anyone wanting to play the fruit machine pays 10 p for a turn.

The winning combinations and amounts are as follows:

Oranges in 3 windows	£1.00
Cherries in 3 windows	£0.50
Oranges in 2 windows and cherries in 1 window	£0.80
Lemons in 3 windows	£0.40

Find the expected gain/loss per turn.

Solution 4.7

$P(\text{oranges in 3 windows}) = (0.1)^3 = 0.001$ (independent events)
$P(\text{cherries in 3 windows}) = (0.2)^3 = 0.008$
$P(\text{oranges in 2 and cherries in 1}) = 3(0.1)^2(0.2) = 0.006$
$P(\text{lemons in 3 windows}) = (0.4)^3 = 0.064$

Therefore

$$P(\text{combination will not win a prize}) = 1 - (0.001 + 0.008 + 0.006 + 0.064) = 0.921$$

Let X be the r.v. 'the amount, in pence, gained per turn'.

Now the amount paid out by the fruit machine could be 100 p, 80 p, 50 p, 40 p or 0 p.

So considering the initial payment of 10 p for a turn, X can take the values 90, 70, 40, 30, −10.

The probability distribution for X is

x	90	70	40	30	−10
$P(X = x)$	0.001	0.006	0.008	0.064	0.921

Now $E(X) = \sum_{\text{all } x} xP(X = x)$

$$= 90(0.001) + 70(0.006) + 40(0.008) + 30(0.064) + (-10)(0.921)$$

$$= -6.46$$

So, the expected *loss* per turn is 6.46 p.

Example 4.8 Three dice are thrown. If a 1 or a 6 turns up, you will be paid 1 p, but if neither a 1 nor a 6 turns up, you will pay 5 p. How much would you expect to gain or lose in 9 games?

You are now given the opportunity to change the rule for payment when a 1 or a 6 appears. To make the game worthwhile to yourself, what is the minimum amount in everyday currency that you would suggest?

Solution 4.8 $P(1 \text{ or } 6 \text{ on die}) = \frac{2}{6} = \frac{1}{3}$.

If three dice are thrown, $P(\text{neither a 1 nor a 6 on all three}) = \left(\frac{2}{3}\right)^3 = \frac{8}{27}$.

So $P(\text{a 1 or a 6 turns up}) = 1 - \frac{8}{27} = \frac{19}{27}$.

Let X be the r.v. 'the number of pence won in a game'. Then X can assume the values −5 and 1 only.

Now $P(X = -5) = P(\text{neither a 1 nor a 6}) = \frac{8}{27}$

$P(X = 1) = P(\text{a 1 or a 6}) = \frac{19}{27}$

The probability distribution for X is

x	−5	1
$P(X = x)$	$\frac{8}{27}$	$\frac{19}{27}$

So
$$E(X) = \sum_{\text{all } x} xP(X = x)$$

$$= (-5)\left(\frac{8}{27}\right) + (1)\left(\frac{19}{27}\right)$$

$$= -\frac{7}{9}$$

Therefore the expected *loss* after one game is $\frac{7}{9}$p,

so, after 9 games, the expected loss is 7 p.

If we change the rule for payment to y pence when a 1 or a 6 turns up then the probability distribution becomes

x	-5	y
$P(X = x)$	$\frac{8}{27}$	$\frac{19}{27}$

We now have
$$E(X) = (-5)\left(\frac{8}{27}\right) + (y)\left(\frac{19}{27}\right)$$
$$= \frac{-40 + 19y}{27}$$

To make the game worthwhile, we require $E(X) > 0$,

so
$$\frac{-40 + 19y}{27} > 0$$

i.e.
$$-40 + 19y > 0$$
$$19y > 40$$
$$y > 2.105\ldots$$

Therefore the minimum amount we should be paid, in everyday currency, is 3 p.

Example 4.9 A bag contains 3 red balls and 1 blue ball. A second bag contains 1 red ball and 1 blue ball. A ball is picked out of each bag and is then placed in the other bag. What is the expected number of red balls in the first bag?

Solution 4.9 Assume that the balls are taken from each bag simultaneously.

If a red ball is picked from each bag and placed in the other then the number of red balls in the first bag is now 3, etc.

Let X be the r.v. 'the final number of red balls in the first bag'.

Then X can take the values 2, 3 or 4 only. With obvious notation,

$$P(X = 2) = P(\text{red from first bag and blue from second bag})$$
$$= P(R_1B_2)$$
$$= \left(\frac{3}{4}\right)\left(\frac{1}{2}\right)$$
$$= \frac{3}{8}$$
$$P(X = 3) = P(R_1R_2) + P(B_1B_2)$$
$$= \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{2}\right)$$
$$= \frac{1}{2}$$

$$P(X = 4) = P(B_1R_2)$$
$$= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right)$$
$$= \tfrac{1}{8}$$

The probability distribution for X is

x	2	3	4
$P(X = x)$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

So
$$E(X) = \sum_{\text{all } x} xP(X = x)$$
$$= 2\left(\tfrac{3}{8}\right) + 3\left(\tfrac{1}{2}\right) + 4\left(\tfrac{1}{8}\right)$$
$$= 2\tfrac{3}{4}$$

The expected number of red balls in the first bag after the exchange is $2\frac{3}{4}$ balls.

Exercise 4b

1. The probability distribution for the r.v. X is shown in the table:

x	0	1	2	3	4
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$

Find $E(X)$.

2. The r.v. X has p.d.f. $P(X = x)$ for $x = 5, 6, 7, 8, 9$ as defined in the table:

x	5	6	7	8	9
$P(X = x)$	$\frac{3}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{2}{11}$	$\frac{3}{11}$

Find μ.

3. The probability distribution of a r.v. X is as shown in the table:

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

Find (*a*) the value of y, (*b*) $E(X)$.

4. Find the expected number of heads when two fair coins are tossed.

5. Find the expected number of ones when three ordinary fair dice are thrown.

6. A bag contains 5 black counters and 6 red counters. Two counters are drawn, one at a time, and not replaced. Let X be the r.v. 'the number of red counters drawn'. Find $E(X)$.

7. An unbiased tetrahedral die has faces marked 1, 2, 3, 4. If the die lands on the face marked 1, the player has to pay 10 p. If it lands on a face marked with a 2 or a 4, the player wins 5 p and if it lands on a 3, the player wins 3 p. Find the expected gain in one throw.

8. A discrete r.v. X can assume values 10 and 20 only. If $E(X) = 16$, write the p.d.f. of X in table form.

9. The discrete r.v. X can assume values 0, 1, 2 and 3 only. Given $P(X \leqslant 2) = 0.9$, $P(X \leqslant 1) = 0.5$ and $E(X) = 1.4$, find (*a*) $P(X = 1)$, (*b*) $P(X = 0)$.

10. In a game, a player rolls two balls down an inclined plane so that each ball finally settles in one of five slots and scores the number of points allotted to that slot as shown in the diagram below:

2	4	7	4	2

It is possible for both balls to settle in one slot and it may be assumed that each slot is equally likely to accept either ball.
The player's score is the sum of the points scored by each ball.
Draw up a table showing all the possible scores and the probability of each.
If the player pays 10 p for each game and receives back a number of pence equal to his score, calculate the player's expected gain or loss per 50 games. (C Additional)

11. In a game a player tosses three fair coins. He wins £10 if 3 heads occur, £x if 2 heads occur, £3 if 1 head occurs and £2 if no heads occur. Express in terms of x his expected gain from each game.
Given that he pays £4.50 to play each game, calculate
(a) the value of x for which the game is fair,
(b) his expected gain or loss over 100 games if $x = 4.90$. (C Additional)

12. A committee of 3 is to be chosen from 4 girls and 7 boys. Find the expected number of girls on the committee, if the members of the committee are chosen at random.

13. The discrete r.v. X has p.d.f. given by $P(X = x) = kx$ for $x = 1, 2, 3, 4, 5$ where k is constant. Find $E(X)$.

14. In an examination a candidate is given the four answers to four questions but is not told which answer applies to which question. He is asked to write down each of the four answers next to its appropriate question.
(a) Calculate in how many different ways he could write down the four answers.
(b) Explain why it is impossible for him to have just three answers in the correct places and show that there are six ways of getting just two answers in the correct places.
(c) If a candidate guesses at random where the four answers are to go and X is the number of correct guesses he makes, draw up the probability distribution for X in tabular form.
(d) Calculate $E(X)$. (L Additional)

15.

x	0	1	2	3
$P(X = x)$	c	c^2	$c^2 + c$	$3c^2 + 2c$

The above table shows the probability distribution for a random variable X.
Calculate (a) c, (b) $E(X)$. (L Additional)

16. A box contains 9 discs of which 4 are red, 3 are white and 2 are blue. Three discs are to be drawn at random without replacement from the box. Calculate
(a) the probability that the discs, in the order drawn, will be coloured red, white and blue respectively,
(b) the probability that one disc of each colour will be drawn,
(c) the probability that the third disc drawn will be red,
(d) the probability that no red disc will be drawn,
(e) the most probable number of red discs that will be drawn,
(f) the expected number of red discs that will be drawn, and state the probability that this expected number of red discs will be drawn. (JMB)

17. A woman has 3 keys on a ring, just one of which opens the front door. As she approaches the front door she selects one key after another at random without replacement. Draw a tree diagram to illustrate the various selections before she finds the correct key. Use this diagram to calculate the expected number of keys that she will use before opening the door. (L Additional)

18. An urn containing 4 black balls and 8 white balls is used for two experiments. In Experiment 1, two balls are to be drawn at random from the urn, one after the other, without replacement. In Experiment 2, one ball is to be drawn at random from the 12 balls in the urn and replaced before a second ball is drawn at random. Copy and complete the following two tables, which give the probabilities for the different compound events in the two experiments.

		Second ball	
		Black	White
First ball	Black		$\frac{8}{33}$
	White		

Experiment 1

		Second ball	
		Black	White
First ball	Black		
	White	$\frac{2}{9}$	

Experiment 2

For each of the two experiments, calculate the expected number of black balls which will be drawn.

If in Experiment 2, the urn contains b black balls and w white balls, where $b + w = 12$, calculate the expected number of black balls which will be drawn.

(L Additional)

THE EXPECTATION OF ANY FUNCTION OF X, $E[g(X)]$

The definition of expectation can be extended to any function of the random variable such as $10X$, X^2, $(X - 4)$, etc.

In general, if $g(X)$ is any function of the discrete random variable X then

$$E[g(X)] = \sum_{\text{all } x} g(x)P(X = x)$$

For example,

$$E(10X) = \sum 10x\,P(X = x)$$

$$E(X^2) = \sum x^2\,P(X = x)$$

$$E(X - 4) = \sum (x - 4)P(X = x)$$

Example 4.10 The random variable X has p.d.f. $P(X = x)$ for $x = 1, 2, 3$.

x	1	2	3
$P(X = x)$	0.1	0.6	0.3

Calculate (**a**) $E(3)$, (**b**) $E(X)$, (**c**) $E(5X)$, (**d**) $E(5X + 3)$, (**e**) $5E(X) + 3$, (**f**) $E(X^2)$, (**g**) $E(4X^2 - 3)$, (**h**) $4E(X^2) - 3$.

Solution 4.10 We have

x	1	2	3
$5x$	5	10	15
$5x + 3$	8	13	18
x^2	1	4	9
$4x^2 - 3$	1	13	33
$P(X = x)$	0.1	0.6	0.3

Now

$$E[g(X)] = \sum_{\text{all } x} g(x)P(X = x)$$

(a)
$$\begin{aligned} E(3) &= \sum_{\text{all } x} 3P(X = x) \\ &= 3(0.1) + 3(0.6) + 3(0.3) \\ &= 3 \\ \underline{E(3)} &\underline{= 3} \end{aligned}$$

(b)
$$\begin{aligned} E(X) &= \sum_{\text{all } x} xP(X = x) \\ &= 1(0.1) + 2(0.6) + 3(0.3) \\ &= 2.2 \\ \underline{E(X)} &\underline{= 2.2} \end{aligned}$$

(c)
$$\begin{aligned} E(5X) &= \sum_{\text{all } x} 5xP(X = x) \\ &= 5(0.1) + 10(0.6) + 15(0.3) \\ &= 11 \\ \underline{E(5X)} &\underline{= 11} \end{aligned}$$

NOTE: $E(5X) = 5E(X)$

(d)
$$\begin{aligned} E(5X+3) &= \sum_{\text{all } x} (5x+3)P(X = x) \\ &= 8(0.1) + 13(0.6) + 18(0.3) \\ &= 14 \\ \underline{E(5X+3)} &\underline{= 14} \end{aligned}$$

(e)
$$\begin{aligned} 5E(X)+3 &= 5(2.2)+3 \\ &= 14 \\ \underline{5E(X)+3} &\underline{= 14} \end{aligned}$$

NOTE: $E(5X+3) = E(5X) + E(3) = 5E(X) + 3$

(f)
$$\begin{aligned} E(X^2) &= \sum_{\text{all } x} x^2P(X = x) \\ &= 1(0.1) + 4(0.6) + 9(0.3) \\ &= 5.2 \\ \underline{E(X^2)} &\underline{= 5.2} \end{aligned}$$

(g)

$$E(4X^2-3) = \sum_{\text{all } x}(4x^2-3)P(X=x)$$
$$= 1(0.1)+13(0.6)+33(0.3)$$
$$= 17.8$$
$$\underline{E(4X^2-3) = 17.8}$$

(h)

$$4E(X^2)-3 = 4(5.2)-3$$
$$= 17.8$$
$$\underline{4E(X^2)-3 = 17.8}$$

NOTE: $E(4X^2-3) = E(4X^2)-E(3) = 4E(X^2)-3$

In general, the following results hold when X is a discrete random variable.

Result 1 $E(a) = a$, where a is any constant.

Proof:

$$E(a) = \sum_{\text{all } x} aP(X=x)$$
$$= a\sum_{\text{all } x} P(X=x)$$
$$= a \qquad \text{since } \sum_{\text{all } x} P(X=x) = 1$$

Result 2 $E(aX) = aE(X)$, where a is any constant.

Proof:

$$E(aX) = \sum_{\text{all } x} axP(X=x)$$
$$= a\sum_{\text{all } x} xP(X=x)$$
$$= aE(X)$$

Result 3 $E(aX+b) = aE(X)+b$, where a and b are any constants.

Proof:

$$E(aX+b) = \sum_{\text{all } x}(ax+b)P(X=x)$$
$$= \sum_{\text{all } x} axP(X=x)+\sum_{\text{all } x} bP(X=x)$$
$$= aE(X)+b$$

Result 4 $E[f_1(X)+f_2(X)] = E[f_1(X)]+E[f_2(X)]$ where f_1 and f_2 are functions of X.

Proof: $E[f_1(X)+f_2(X)] = \sum_{\text{all } x} [f_1(x)+f_2(x)]P(X=x)$

$$= \sum_{\text{all } x} f_1(x)P(X=x) + \sum_{\text{all } x} f_2(x)P(X=x)$$

$$= E[f_1(X)]+E[f_2(X)]$$

Example 4.11 The discrete random variable X has the probability distribution specified in the following table.

x	-1	0	1	2
$P(X=x)$	0.25	0.10	0.45	0.20

(**a**) Find $P(-1 \leqslant X < 1)$.

(**b**) Find $E(2X+3)$. (L)

Solution 4.11 (a)

$$P(-1 \leqslant X < 1) = P(X=-1)+P(X=0)$$
$$= 0.25+0.10$$
$$= 0.35$$

Therefore $P(-1 \leqslant X < 1) = 0.35.$

(**b**) $E(2X+3) = 2E(X)+3$, so we need to find $E(X)$.

Now $E(X) = \sum_{\text{all } x} xP(X=x)$

$$= (-1)(0.25)+(0)(0.10)+(1)(0.45)+(2)(0.20)$$
$$= 0.6$$

So $E(2X+3) = 2(0.6)+3$

$$= 4.2$$

Therefore $E(2X+3) = 4.2.$

Exercise 4c

1. The discrete r.v. X has p.d.f. $P(X=x)$ for $x = 1, 2, 3$.

x	1	2	3
$P(X=x)$	0.2	0.3	0.5

Find (*a*) $E(X)$, (*b*) $E(X^2)$.

(*c*) Verify that $E(3X-1) = 3E(X)-1$.

(*d*) Verify that $E(2X^2+4) = 2E(X^2)+4$.

2. The discrete r.v. X has p.d.f. $P(X=0) = 0.05$, $P(X=1) = 0.45$, $P(X=2) = 0.5$.

Find (*a*) $\mu = E(X)$, (*b*) $E(X^2)$, (*c*) $E(5X^2+2X-3)$.

3. The discrete r.v. X has p.d.f. given by
$P(X = x) = k$ for $x = 1, 2, 3, 4, 5, 6$.
Find (a) $E(X)$, (b) $E(X^2)$, (c) $E(3X + 4)$, (d) $E(2X^2 + X - 4)$.

4. The discrete r.v. X has p.d.f. given by
$P(X = x) = \dfrac{3x+1}{22}$ for $x = 0, 1, 2, 3$.
Find (a) $E(X)$, (b) $E(X^2)$, (c) $E(3X - 2)$, (d) $E(2X^2 + 4X - 3)$.

5. A roulette wheel is divided into 6 sectors of unequal area, marked with the numbers 1, 2, 3, 4, 5 and 6. The wheel is spun and X is the r.v. 'the number on which the wheel stops'. The probability distribution of X is as follows:

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{16}$

Calculate (a) $E(X)$, (b) $E(X^2)$, (c) $E(3X - 5)$, (d) $E(6X^2)$, (e) $E(6X^2 + 6X - 10)$.

6. The r.v. X has p.d.f. $P(X = x)$ as shown in the table:

x	-2	-1	0	1	c
$P(X = x)$	0.1	0.1	0.3	0.4	0.1

Find the value of c (a) if $E(X) = 0.3$ (b) if $E(X^2) = 1.8$.

7. The discrete random variable X has probability function given by

$$p(x) = \begin{cases} \left(\frac{1}{2}\right)^x & x = 1,2,3,4,5, \\ C & x = 6, \\ 0 & \text{otherwise,} \end{cases}$$

where C is a constant.
Determine the value of C and hence the mode and arithmetic mean of X. (L)

8. The probability distribution of a discrete random variable X is given by

$$P(X = r) = kr, \qquad r = 1,2,3,\dots,n,$$

where k is a constant.
Show that

$$k = \frac{2}{n(n+1)}$$

and find, in terms of n, the mean of X. (JMB)

VARIANCE, Var (X)

Experimental approach

For a frequency distribution with mean $\bar{x}$, the variance, s^2 is given by

$$s^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}. \quad \text{This can also be written} \quad s^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2.$$

Theoretical approach

For a discrete random variable X, with $E(X) = \mu$, the variance is defined as follows:

> The variance of X, written Var(X), is given by
>
> $$\text{Var}(X) = E(X - \mu)^2$$

Alternatively,

$$\begin{aligned} \text{Var}(X) &= E(X - \mu)^2 \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

So we have $\boxed{\text{Var}(X) = E(X^2) - \mu^2}$

NOTE: $\mu = E(X)$, so $\mu^2 = [E(X)]^2$. We write $[E(X)]^2$ as $E^2(X)$ in a similar way to the notation used in trigonometry where $(\sin A)^2$ is written $\sin^2 A$.

So we have $\boxed{\text{Var}(X) = E(X^2) - E^2(X)}$

Example 4.12 The r.v. X has probability distribution as shown in the table:

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

Find

(a) $\mu = E(X)$,

(b) $\text{Var}(X)$, using the formula $\text{Var}(X) = E(X - \mu)^2$,

(c) $E(X^2)$,

(d) $\text{Var}(X)$, using the formula $\text{Var}(X) = E(X^2) - \mu^2$.

Solution 4.12 (a) By symmetry, $\mu = E(X) = 3$.

(b)
$$\begin{aligned} E(X-\mu)^2 &= E(X-3)^2 \\ &= \sum_{\text{all } x} (x-3)^2 P(X = x) \end{aligned}$$

x	1	2	3	4	5
$(x - 3)$	-2	-1	0	1	2
$(x - 3)^2$	4	1	0	1	4
$P(X = x)$	0.1	0.3	0.2	0.3	0.1

So
$$\begin{aligned} E(X-3)^2 &= 4(0.1) + 1(0.3) + 0(0.2) + 1(0.3) + 4(0.1) \\ &= 1.4 \end{aligned}$$

Therefore $\text{Var}(X) = E(X - \mu)^2 = 1.4$.

(c)
$$\begin{aligned} E(X^2) &= \sum_{\text{all } x} x^2 P(X = x) \\ &= 1(0.1) + 4(0.3) + 9(0.2) + 16(0.3) + 25(0.1) \\ &= 10.4 \end{aligned}$$

So $E(X^2) = 10.4$.

(**d**) Now
$$\text{Var}(X) = E(X^2) - \mu^2 = 10.4 - 9 = 1.4$$

Therefore Var(X) = 1.4, as before.

Example 4.13 Two discs are drawn, without replacement, from a box containing 3 red discs and 4 white discs. The discs are drawn at random. If X is the r.v. 'the number of red discs drawn', find

(**a**) the expected number of red discs

(**b**) the standard deviation of X.

Solution 4.13 X is the r.v. 'the number of red discs drawn'.

Now X can take the values 0, 1, 2 only. We have

$$P(X=0) = P(W_1W_2) = \left(\tfrac{4}{7}\right)\left(\tfrac{3}{6}\right) = \tfrac{12}{42} = \tfrac{2}{7}$$

$$P(X=1) = P(W_1R_2) + P(R_1W_2) = \left(\tfrac{4}{7}\right)\left(\tfrac{3}{6}\right) + \left(\tfrac{3}{7}\right)\left(\tfrac{4}{6}\right) = \tfrac{24}{42} = \tfrac{4}{7}$$

$$P(X=2) = P(R_1R_2) = \left(\tfrac{3}{7}\right)\left(\tfrac{2}{6}\right) = \tfrac{6}{42} = \tfrac{1}{7}$$

The probability distribution for X is as follows:

x	0	1	2
$P(X=x)$	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

(**a**) Now
$$E(X) = \sum_{\text{all } x} xP(X=x) = 0\left(\tfrac{2}{7}\right) + 1\left(\tfrac{4}{7}\right) + 2\left(\tfrac{1}{7}\right) = \tfrac{6}{7}$$

So the expected number of red discs is $\frac{6}{7}$.

(**b**) Standard deviation of $X = \sqrt{\text{Var}(X)}$.

Now
$$\text{Var}(X) = E(X^2) - E^2(X)$$

We have
$$E(X^2) = \sum_{\text{all } x} x^2P(X=x) = 0\left(\tfrac{2}{7}\right) + 1\left(\tfrac{4}{7}\right) + 4\left(\tfrac{1}{7}\right) = \tfrac{8}{7}$$

So
$$\text{Var}(X) = \tfrac{8}{7} - \left(\tfrac{6}{7}\right)^2 = \tfrac{20}{49}$$

Therefore the standard deviation of $X = \sqrt{20/49} = 0.639$ (3 d.p.).

The following results are useful

Result 1 — Var$(a) = 0$ where a is any constant.

Proof:

$$\begin{aligned}\text{Var}(a) &= E(a^2) - E^2(a)\\ &= a^2 - a^2\\ &= 0\end{aligned}$$

NOTE: this is as expected, since a constant does not vary.

Result 2 — Var$(aX) = a^2$Var(X) where a is any constant.

Proof:

$$\begin{aligned}\text{Var}(aX) &= E(aX)^2 - E^2(aX)\\ &= a^2E(X^2) - a^2E^2(X)\\ &= a^2[E(X^2) - E^2(X)]\\ &= a^2\text{Var}(X)\end{aligned}$$

Result 3 — Var$(aX + b) = a^2$Var(X) where a and b are any constants.

Proof:

$$\begin{aligned}\text{Var}(aX+b) &= E(aX+b)^2 - E^2(aX+b)\\ &= E(a^2X^2 + 2abX + b^2) - [aE(X)+b]^2\\ &= a^2E(X^2) + 2abE(X) + b^2 - a^2E^2(X) - 2abE(X) - b^2\\ &= a^2E(X^2) - a^2E^2(X)\\ &= a^2[E(X^2) - E^2(X)]\\ &= a^2\text{Var}(X)\end{aligned}$$

Example 4.14 The discrete random variable X has probability distribution as shown in the table. Find Var$(2X + 3)$.

x	10	20	30
$P(X = x)$	0.1	0.6	0.3

Solution 4.14 Now $\text{Var}(2X+3) = 4\,\text{Var}(X)$ (Result 3)

We will need to find $\text{Var}(X) = E(X^2) - E^2(X)$.

$$E(X) = \sum_{\text{all } x} xP(X=x)$$
$$= 10(0.1) + 20(0.6) + 30(0.3)$$
$$= 22$$
$$E(X^2) = \sum_{\text{all } x} x^2P(X=x)$$
$$= 100(0.1) + 400(0.6) + 900(0.3)$$
$$= 520$$
$$\text{Var}(X) = E(X^2) - E^2(X)$$
$$= 520 - 22^2$$
$$= 36$$
$$\text{Var}(2X+3) = 4\,\text{Var}(X)$$
$$= 144$$

Therefore $\text{Var}(2X+3) = 144$.

Exercise 4d

1. Find Var(X) for each of the following probability distributions:

(a)

x	-3	-2	0	2	3
$P(X=x)$	0.3	0.3	0.2	0.1	0.1

(b)

x	1	3	5	7	9
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$

(c)

x	0	2	5	6
$P(X=x)$	0.11	0.35	0.46	0.08

2. If X is the r.v. 'the number on a biased die', and the p.d.f. of X is as shown,

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

Find (a) the value of y, (b) $E(X)$, (c) $E(X^2)$, (d) Var(X), (e) Var(4X).

3. If X is the r.v. 'the sum of the scores on two tetrahedral dice', where the 'score' is the number on which the die lands, find (a) $E(X)$, (b) Var(X), (c) Var(2X), (d) Var(2X + 3).

4. A team of 3 is to be chosen from 4 boys and 5 girls. If X is the r.v. 'the number of girls in the team', find (a) $E(X)$, (b) $E(X^2)$, (c) Var(X).

5. Two discs are drawn without replacement from a box containing 3 red and 4 white discs. If X is the r.v. 'the number of white discs drawn', construct a probability distribution table. Find (a) $E(X)$, (b) $E(X^2)$, (c) Var(X), (d) Var(3X − 4).

6. For the following probability distribution find (a) $\mu = E(X)$, (b) $E(X^2)$, (c) $E(X-\mu)^2$. Verify that

$$E(X-\mu)^2 = E(X^2) - \mu^2$$

x	-3	-2	1
$P(X=x)$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{2}$

7. Ten identically shaped discs are in a bag; two of them are black, the rest white. Discs are drawn at random from the bag in turn and not replaced.
Let X be the number of discs drawn up to and including the first black one.
List the values of X and the associated theoretical probabilities.
Calculate the mean value of X and its standard deviation. What is the most likely value of X?
If instead each disc is replaced before the next is drawn, construct a similar list of values and point out the chief differences between the two lists.

8. The discrete r.v. X has p.d.f.

$$P(X = x) = k|x|$$

where x takes the values $-3, -2, -1, 0, 1, 2, 3$. Find (*a*) the value of the constant k, (*b*) $E(X)$, (*c*) $E(X^2)$, (*d*) the standard deviation of X.

9. The random variable X takes integer values only and has p.d.f.

$P(X = x) = kx \qquad x = 1, 2, 3, 4, 5$

$P(X = x) = k(10 - x) \qquad x = 6, 7, 8, 9$

Find (*a*) the value of the constant k, (*b*) $E(X)$, (*c*) $\text{Var}(X)$, (*d*) $E(2X - 3)$, (*e*) $\text{Var}(2X - 3)$.

10. (*a*) In a game a player pays £5 to toss three fair coins. Depending on the number of tails he obtains he receives a sum of money as shown in the table below.

Number of tails	3	2	1	0
Sum received	£10	£6	£3	£1

Calculate the player's expected gain or loss over 12 games.
(*b*) A variable X has a probability distribution shown in the table below.

Value of X	1	2	5	10
Probability	0.5	0.3	p	q

Given that X can only take the values 1, 2, 5 or 10, and that $E(X) = 2.5$, calculate
(i) the value of p and of q,
(ii) the variance of X.
In a fairground game, a player rolls discs on to a board containing squares, each of which bears one of the numbers 1, 2, 5 or 10. If a disc falls entirely within a square, the player receives the same number of pence as the number in the square; if it does not, the player does not receive anything. The probability that a player will receive money from any given roll is $\frac{1}{4}$. If a player does receive money, the probabilities of receiving 1, 2, 5 or 10 pence are the same as those connected with the values of X above. How many discs should a player be allowed to roll for 5 p, if the game is to be fair?
(C Additional)

11. (*a*) A man takes part in a game in which he throws two fair dice and scores the sum of the two numbers shown. The rewards for the scores are given in the following table.

Score	12	10	7	5	Any other score
Reward (£)	16	6	3	5	0

Calculate the expected reward for a throw of the two dice.
(*b*) A bag contains five identical discs, two of which are marked with the letter A and three with the letter B. The discs are randomly drawn, one at a time without replacement, until both discs marked A are obtained. Show that the probability that 3 draws are required is $\frac{2}{10}$.
Given that X denotes the number of draws required to obtain both discs marked A, copy and complete the following table.

Value of X	2	3	4	5
Probability of X		$\frac{2}{10}$		

Evaluate (i) $E(X)$, (ii) $E(X^2)$, (iii) the variance of X. (C Additional)

THE CUMULATIVE DISTRIBUTION FUNCTION

Given a frequency distribution, the corresponding cumulative frequencies are obtained by summing all the frequencies up to a particular value. In the same way, if X is a discrete random variable, the corresponding cumulative probabilities are obtained by summing all the probabilities up to a particular value.

If X is a discrete random variable with p.d.f. $P(X = x)$ for $x = x_1$, $x_2, \ldots, x_n$, then the cumulative distribution function is given by $F(t)$ where

$$F(t) = P(X \leqslant t)$$
$$= \sum_{x=x_1}^{t} P(X = x) \qquad t = x_1, x_2, \ldots, x_n$$

The cumulative distribution function is sometimes known just as the **distribution function**.

Example 4.15 Find the cumulative distribution function for the r.v. X where X is 'the score on an unbiased die'.

Solution 4.15 The probability distribution for X is shown in the table:

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

$$F(1) = P(X \leqslant 1) = \tfrac{1}{6}$$
$$F(2) = P(X \leqslant 2) = P(X = 1) + P(X = 2) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{2}{6}$$
$$F(3) = P(X \leqslant 3) = \tfrac{3}{6}$$
$$F(4) = P(X \leqslant 4) = \tfrac{4}{6}$$
$$F(5) = P(X \leqslant 5) = \tfrac{5}{6}$$
$$F(6) = P(X \leqslant 6) = \tfrac{6}{6}$$

Therefore $F(t) = \frac{t}{6}$ for $t = 1, 2, 3, \ldots, 6$.

Although we work with the variable t, we write the final answer in terms of x, so

$$\underline{F(x) = \tfrac{x}{6}, \qquad x = 1, 2, \ldots, 6.}$$

NOTE: $F(6) = \frac{6}{6} = 1$, as expected.

Example 4.16 The probability distribution for the r.v. X is shown in the table. Construct the cumulative distribution table.

x	0	1	2	3	4	5	6
$P(X = x)$	0.03	0.04	0.06	0.12	0.4	0.15	0.2

Solution 4.16 Now $$F(t) = \sum_{x=0}^{t} P(X = x) \qquad t = 0, 1, 2, \ldots, 6$$

So
$$F(0) = P(X \leqslant 0) = 0.03$$
$$F(1) = P(X \leqslant 1) = 0.03 + 0.04 = 0.07$$
$$F(2) = P(X \leqslant 2) = 0.03 + 0.04 + 0.06 = 0.13$$

and so on.

The cumulative distribution table is:

x	0	1	2	3	4	5	6
$F(x)$	0.03	0.07	0.13	0.25	0.65	0.8	1

NOTE: it is not possible to write a formula for the cumulative distribution function in this example.

Example 4.17 For a discrete r.v. X the cumulative distribution function $F(x)$ is as shown:

x	1	2	3	4	5
$F(x)$	0.2	0.32	0.67	0.9	1

Find **(a)** $P(X = 3)$, **(b)** $P(X > 2)$.

Solution 4.17 **(a)** From the table,

$$F(3) = P(X \leqslant 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.67$$
$$F(2) = P(X \leqslant 2) = P(X = 1) + P(X = 2) = 0.32$$

Now $$P(X = 3) = F(3) - F(2)$$

Therefore $$P(X = 3) = 0.67 - 0.32$$
$$\underline{= 0.35}$$

(b) $$P(X > 2) = 1 - P(X \leqslant 2)$$
$$= 1 - F(2)$$
$$= 1 - 0.32$$
$$= 0.68$$

<u>So $P(X = 3) = 0.35$ and $P(X > 2) = 0.68$.</u>

Exercise 4e

1. Construct the cumulative distribution tables for the following discrete random variables:
(*a*) the number of sixes obtained when two ordinary dice are thrown,
(*b*) the smaller number when two ordinary dice are thrown,
(*c*) the number of heads when three fair coins are tossed.

2. The probability distribution for the r.v. Y is shown in the table:

y	0.1	0.2	0.3	0.4	0.5
$P(Y = y)$	0.05	0.25	0.3	0.15	0.25

Construct the cumulative distribution table.

3. For a discrete r.v. R the cumulative distribution function $F(r)$ is as shown in the table:

r	1	2	3	4
$F(r)$	0.13	0.54	0.75	1

Find (*a*) $P(R = 2)$, (*b*) $P(R > 1)$, (*c*) $P(R \geqslant 3)$, (*d*) $P(R < 2)$, (*e*) $E(R)$.

4. For the discrete r.v. X the cumulative distribution function $F(x)$ is as shown:

x	3	4	5	6	7
$F(x)$	0.01	0.23	0.64	0.86	1

Construct the probability distribution of X, and find $\text{Var}(X)$.

5. For a discrete r.v. X the cumulative distribution function is given by $F(x) = \frac{x^2}{9}$ for $x = 1, 2, 3$. Find (*a*) $F(2)$, (*b*) $P(X = 2)$. (*c*) Write out the probability distribution of X. (*d*) Find $E(2X - 3)$.

6. For a discrete r.v. X the cumulative distribution function is given by $F(x) = kx$, $x = 1, 2, 3$. Find (*a*) the value of the constant k, (*b*) $P(X < 3)$, (*c*) the probability distribution of X, (*d*) the standard deviation of X.

7. The discrete r.v. X has distribution function $F(x)$ where
$F(x) = 1 - \left(1 - \frac{1}{4}x\right)^x$ for $x = 1, 2, 3, 4$
(*a*) Show that $F(3) = \frac{63}{64}$ and $F(2) = \frac{3}{4}$.
(*b*) Obtain the probability distribution of X.
(*c*) Find $E(X)$ and $\text{Var}(X)$.
(*d*) Find $P(X > E(X))$.

TWO INDEPENDENT RANDOM VARIABLES

If X and Y are any two random variables, then

$$E(X + Y) = E(X) + E(Y)$$

If X and Y are *independent* random variables, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Example 4.18 X is the r.v. 'the score on a tetrahedral die', Y is the r.v. 'the number of heads obtained when two coins are tossed'.

(a) Obtain the probability distributions of X and of Y.

(b) Find $E(X)$ and $E(Y)$.

(c) Find $\text{Var}(X)$ and $\text{Var}(Y)$.

(d) Obtain the probability distribution for the r.v. $X + Y$.

(e) Find $E(X + Y)$ and $\text{Var}(X + Y)$ using the probability distribution for $X + Y$. Comment on your results.

Solution 4.18 (a) The probability distributions are as follows:

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

y	0	1	2
$P(Y = y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(b) By symmetry, $\underline{E(X) = 2\frac{1}{2}}$ $\qquad$ $\underline{E(Y) = 1}$

(c)
$$\begin{aligned} E(X^2) &= \sum_{\text{all } x} x^2 P(X = x) \\ &= 1\left(\tfrac{1}{4}\right) + 4\left(\tfrac{1}{4}\right) \\ &\quad + 9\left(\tfrac{1}{4}\right) + 16\left(\tfrac{1}{4}\right) \\ &= 7\tfrac{1}{2} \\ \text{Var}(X) &= E(X^2) - E^2(X) \\ &= 7\tfrac{1}{2} - 6\tfrac{1}{4} \\ &= 1\tfrac{1}{4} \end{aligned}$$

So $\underline{\text{Var}(X) = 1\frac{1}{4}}$.

$$\begin{aligned} E(Y^2) &= \sum_{\text{all } y} y^2 P(Y = y) \\ &= 0\left(\tfrac{1}{4}\right) + 1\left(\tfrac{1}{2}\right) \\ &\quad + 4\left(\tfrac{1}{4}\right) \\ &= 1\tfrac{1}{2} \\ \text{Var}(Y) &= E(Y^2) - E^2(Y) \\ &= 1\tfrac{1}{2} - 1 \\ &= \tfrac{1}{2} \end{aligned}$$

$\underline{\text{Var}(Y) = \frac{1}{2}}$.

(d) Consider the r.v. $X + Y$, which can take values 1, 2, 3, 4, 5 and 6.

$$P(X+Y=1) = P(1 \text{ on die, } 0 \text{ heads}) = \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) = \tfrac{1}{16}$$

$$\begin{aligned} P(X+Y=2) &= P(2 \text{ on die, } 0 \text{ heads}) + P(1 \text{ on die, } 1 \text{ head}) \\ &= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right) = \tfrac{3}{16} \end{aligned}$$

$$\begin{aligned} P(X+Y=3) &= P(3 \text{ on die, } 0 \text{ heads}) + P(2 \text{ on die, } 1 \text{ head}) \\ &\quad + P(1 \text{ on die, } 2 \text{ heads}) \\ &= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) = \tfrac{4}{16} \end{aligned}$$

$$\begin{aligned} P(X+Y=4) &= P(4 \text{ on die, } 0 \text{ heads}) + P(3 \text{ on die, } 1 \text{ head}) \\ &\quad + P(2 \text{ on die, } 2 \text{ heads}) \\ &= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) \\ &= \tfrac{4}{16} \end{aligned}$$

$$\begin{aligned} P(X+Y=5) &= P(4 \text{ on die, } 1 \text{ head}) + P(3 \text{ on die, } 2 \text{ heads}) \\ &= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right) + \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) \\ &= \tfrac{3}{16} \end{aligned}$$

$$\begin{aligned} P(X+Y=6) &= P(4 \text{ on die, } 2 \text{ heads}) \\ &= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) \\ &= \tfrac{1}{16} \end{aligned}$$

The probability distribution is as follows:

$x+y$	1	2	3	4	5	6
$P(X+Y=x+y)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

(e) By symmetry $E(X+Y) = 3\frac{1}{2}$

But from (b) $E(X)+E(Y) = 2\frac{1}{2}+1 = 3\frac{1}{2}$

Therefore $\underline{E(X+Y) = E(X)+E(Y).}$

Now

$$\text{Var}(X+Y) = E[(X+Y)^2] - E^2(X+Y)$$

$$\begin{aligned} E[(X+Y)^2] &= 1\left(\frac{1}{16}\right)+4\left(\frac{3}{16}\right)+9\left(\frac{4}{16}\right)+16\left(\frac{4}{16}\right)+25\left(\frac{3}{16}\right)+36\left(\frac{1}{16}\right) \\ &= \frac{224}{16} \\ &= 14 \end{aligned}$$

$$\begin{aligned} \text{Var}(X+Y) &= 14-\left(3\tfrac{1}{2}\right)^2 \\ &= 1\tfrac{3}{4} \end{aligned}$$

Therefore $\underline{\text{Var}(X+Y) = 1\frac{3}{4}}.$

Now $\text{Var}(X)+\text{Var}(Y) = 1\frac{1}{4}+1\frac{1}{2} = 1\frac{3}{4}.$

So $\underline{\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y).}$

In this example, the variables X and Y are independent.

In general, for random variables X and Y and constants a and b,

$$E(aX+bY) = aE(X)+bE(Y)$$

If X and Y are *independent*, then

$$\text{Var}(aX+bY) = a^2\text{Var}(X)+b^2\text{Var}(Y)$$

Example 4.19 X and Y are independent discrete random variables such that $E(X)=10$, $\text{Var}(X)=2$, $E(Y)=8$, $\text{Var}(Y)=3$.

Find (a) $E(5X+4Y)$, (b) $\text{Var}(5X+4Y)$, (c) $\text{Var}\left(\frac{1}{2}X+Y\right)$.

Solution 4.19 (a)

$$\begin{aligned} E(5X+4Y) &= 5E(X)+4E(Y) \\ &= 5(10)+4(8) \\ &= \underline{82.} \end{aligned}$$

(b)

$$\begin{aligned} \text{Var}(5X+4Y) &= 25\text{Var}(X)+16\text{Var}(Y) \\ &= 25(2)+16(3) \\ &= \underline{98.} \end{aligned}$$

(c)
$$\begin{aligned}\text{Var}\left(\tfrac{1}{2}X + Y\right) &= \tfrac{1}{4}\text{Var}(X) + \text{Var}(Y)\\ &= \tfrac{1}{4}(2) + 3\\ &= \underline{3.5.}\end{aligned}$$

An important application of the general results occurs when the constant b is negative. For example, if $a = 1$ and $b = -1$,

$$aX + bY = X + (-1)Y = X - Y$$

So
$$\begin{aligned}E(X - Y) &= E(X + (-1)Y)\\ &= E(X) + (-1)E(Y)\\ &= E(X) - E(Y)\\ \text{Var}(X - Y) &= \text{Var}(X + (-1)Y)\\ &= 1^2\text{Var}(X) + (-1)^2\text{Var}(Y)\\ &= \text{Var}(X) + \text{Var}(Y)\end{aligned}$$

Remember that, for X and Y *independent*,

$$E(X - Y) = E(X) - E(Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

Also note that

$$E(aX - bY) = aE(X) - bE(Y)$$

$$\text{Var}(aX - bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Example 4.20 The random variables X and Y are independent and $E(X) = 2$, $\text{Var}(X) = 0.5$, $E(Y) = 5$, $\text{Var}(Y) = 2$.

Find **(a)** $E(4X - 3Y)$, **(b)** $\text{Var}(4X - 3Y)$.

Solution 4.20 (a)
$$\begin{aligned}E(4X - 3Y) &= 4E(X) - 3E(Y)\\ &= 4(2) - 3(5)\\ &= \underline{-7.}\end{aligned}$$

(b)
$$\begin{aligned}\text{Var}(4X - 3Y) &= 4^2\text{Var}(X) + 3^2\text{Var}(Y)\\ &= 16\text{Var}(X) + 9\text{Var}(Y)\\ &= 16(0.5) + 9(2)\\ &= \underline{26.}\end{aligned}$$

Example 4.21 The table gives the joint probability distribution of two random variables X and Y:

	$X = 0$	$X = 1$
$Y = 1$	0.2	0.4
$Y = 2$	0.3	0.1

Calculate (**a**) $E(X)$, (**b**) $E(Y)$, (**c**) $E(X+Y)$.

Solution 4.21 Consider the r.v. X

$$\begin{aligned} P(X=0) &= P(X=0)P(Y=1)+P(X=0)P(Y=2) \\ &= 0.2+0.3 \\ &= 0.5 \\ P(X=1) &= P(X=1)P(Y=1)+P(X=1)P(Y=2) \\ &= 0.4+0.1 \\ &= 0.5 \end{aligned}$$

The probability distribution for X is

x	0	1
$P(X = x)$	0.5	0.5

By symmetry $E(X) = \frac{1}{2}$.

Consider the r.v. Y

$$\begin{aligned} P(Y=1) &= P(Y=1)P(X=0)+P(Y=1)P(X=1) \\ &= 0.2+0.4 \\ &= 0.6 \\ P(Y=2) &= P(Y=2)P(X=0)+P(Y=2)P(X=1) \\ &= 0.3+0.1 \\ &= 0.4 \end{aligned}$$

The probability distribution for Y is

y	1	2
$P(Y = y)$	0.6	0.4

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(Y=y) \\ &= 1(0.6)+2(0.4) \\ &= 1.4 \end{aligned}$$

Therefore $E(Y) = 1.4$.

Now $E(X+Y) = E(X) + E(Y)$

$= 0.5 + 1.4$

$= 1.9$

Therefore $E(X+Y) = 1.9.$

THE DISTRIBUTION OF $X_1 + X_2$

We now consider the distribution of $X_1 + X_2$, where X_1, X_2 are two independent observations *from the same distribution* X.

Now
$$\begin{aligned} E(X_1+X_2) &= E(X_1)+E(X_2) \\ &= E(X)+E(X) \\ &= 2E(X) \end{aligned}$$

and
$$\begin{aligned} \mathrm{Var}(X_1+X_2) &= \mathrm{Var}(X_1)+\mathrm{Var}(X_2) \\ &= \mathrm{Var}(X)+\mathrm{Var}(X) \\ &= 2\mathrm{Var}(X) \end{aligned}$$

For the distribution $X_1 + X_2$, where X_1 and X_2 are *independent* observations from the distribution X

$$E(X_1+X_2) = 2E(X)$$

$$\mathrm{Var}(X_1+X_2) = 2\mathrm{Var}(X)$$

For n independent observations

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

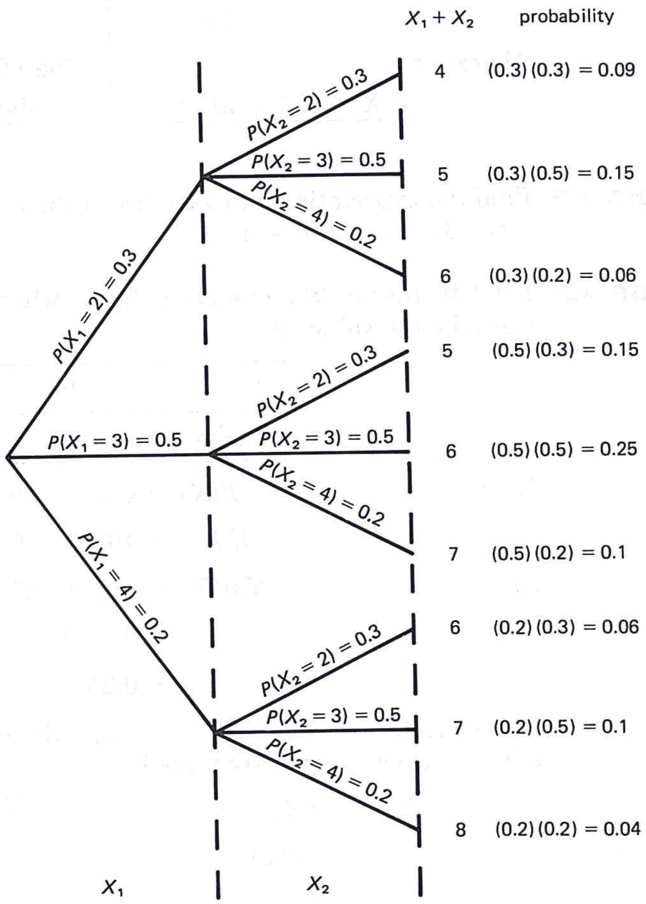
$$\mathrm{Var}(X_1+X_2+\ldots+X_n) = n\mathrm{Var}(X)$$

Example 4.22 The random variable X is such that $E(X) = 2.9$, and the standard deviation of X is 0.7. Its p.d.f. is as shown:

x	2	3	4
$P(X = x)$	0.3	0.5	0.2

Two independent observations are made from X. Construct the probability distribution for $X_1 + X_2$ and find the expectation and variance. Verify that $E(X_1+X_2) = 2E(X)$ and $\mathrm{Var}(X_1+X_2) = 2\mathrm{Var}(X)$.

Solution 4.22 Consider the distribution of X_1+X_2. To show the possible outcomes it is useful to draw a probability tree.



Now $P(X_1+X_2=4) = 0.09$, $P(X_1+X_2=5) = 0.15+0.15 = 0.3$, and so on. We see that X_1+X_2 can take values 4, 5, 6, 7, 8 where

x_1+x_2	4	5	6	7	8
$P(X_1+X_2=x_1+x_2)$	0.09	0.3	0.37	0.2	0.04

$$E(X_1+X_2) = 4(0.09)+5(0.3)+6(0.37)+7(0.2)+8(0.04)$$
$$= 5.8$$
$$\text{Var}(X_1+X_2) = 16(0.09)+25(0.3)+36(0.37)+49(0.2)$$
$$+64(0.04)-5.8^2$$
$$= 0.98$$

Therefore $E(X_1+X_2) = 5.8$ and $\text{Var}(X_1+X_2) = 0.98.$

Now $E(X_1+X_2) = 5.8$

and $2E(X) = 2(2.9) = 5.8$

Therefore $\underline{E(X_1+X_2) = 2E(X).}$

$\mathrm{Var}(X_1+X_2) = 0.98$

$2\mathrm{Var}(X) = 2(0.7)^2 = 0.98$

Therefore $\underline{\mathrm{Var}(X_1+X_2) = 2\mathrm{Var}(X).}$

Example 4.23 Find the expectation and variance of the number of heads obtained when 6 coins are tossed.

Solution 4.23 Let X be the r.v. 'the number of heads when a coin is tossed'. Then X can take the values 0, 1.

x	0	1
$P(X = x)$	0.5	0.5

Now
$$E(X) = 0.5 \quad \text{(by symmetry)}$$
$$E(X^2) = 1(0.5) = 0.5$$
so
$$\mathrm{Var}(X) = E(X^2) - E^2(X) = 0.5 - 0.5^2 = 0.25$$

Now consider $Y = X_1 + X_2 + \ldots + X_6$ where Y is the r.v. 'the number of heads when 6 coins are tossed'.

$$E(Y) = 6E(X) = 6(0.5) = 3$$

$$\mathrm{Var}(Y) = 6\mathrm{Var}(X) = 6(0.25) = 1.5$$

<u>So the expected number of heads is 3, and the variance is 1.5.</u>

Example 4.24 A random variable X has p.d.f. $P(X = x) = kx$ for $x = 1, 2, 3, 4$. Two independent observations of X are made. Let these values be X_1 and X_2.
Find (**a**) $P(X_1 = X_2)$, (**b**) $P(X_1 > X_2)$, (**c**) $P(X_1 = X_2 = 4 \mid X_1 = X_2)$.

Solution 4.24

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

Now
$$\sum_{all\,x} P(X = x) = 1$$
so
$$k + 2k + 3k + 4k = 1$$
$$10k = 1$$
$$k = 0.1$$

(a) $$\begin{aligned} P(X_1 = X_2) &= P((X_1 = 1) \cap (X_2 = 1)) + P((X_1 = 2) \cap (X_2 = 2)) \\ &\quad + P((X_1 = 3) \cap (X_2 = 3)) + P((X_1 = 4) \cap (X_2 = 4)) \\ &= (k)(k) + (2k)(2k) + (3k)(3k) + (4k)(4k) \\ &= k^2 + 4k^2 + 9k^2 + 16k^2 \\ &= 30k^2 \\ &= 30(0.1)^2 \\ &= 0.3 \end{aligned}$$

Therefore $P(X_1 = X_2) = 0.3.$

(b) $$\begin{aligned} P(X_1 > X_2) &= P((X_1 = 2) \cap (X_2 = 1)) + P((X_1 = 3) \cap (X_2 = 1 \text{ or } 2)) \\ &\quad + P((X_1 = 4) \cap (X_2 = 1, 2 \text{ or } 3)) \\ &= (2k)(k) + (3k)(k + 2k) + 4k(k + 2k + 3k) \\ &= 2k^2 + 9k^2 + 24k^2 \\ &= 35k^2 \\ &= 35(0.1)^2 \\ &= 0.35 \end{aligned}$$

Therefore $P(X_1 > X_2) = 0.35.$

(c) $$\begin{aligned} P((X_1 = X_2 = 4) \mid (X_1 = X_2)) &= \frac{P((X_1 = 4) \cap (X_2 = 4))}{P(X_1 = X_2)} \\ &= \frac{16k^2}{30k^2} \\ &= \frac{8}{15} \end{aligned}$$

Therefore $P((X_1 = X_2 = 4) \mid (X_1 = X_2)) = \frac{8}{15}.$

COMPARING THE DISTRIBUTIONS OF $2X$ AND $X_1 + X_2$

Confusion often arises over the different random variables $2X$ and $X_1 + X_2$, where X_1, X_2 are two independent observations of X. We will see from the following example that the distributions of the two random variables are very different.

Example 4.25 When a tetrahedral die is thrown, the number on the face on which it lands, X, has probability distribution:

x	1	2	3	4
$P(X = x)$	0.25	0.25	0.25	0.25

$E(X) = 2.5$ and $\text{Var}(X) = 1.25.$

(a) Find the p.d.f. of D, the r.v. 'double the number on which the die lands', where $D = 2X$, and find $E(D)$ and $\mathrm{Var}(D)$.

(b) Find the p.d.f. of S, the r.v. 'the sum of the two numbers when the die is thrown twice', where $S = X_1 + X_2$, and find $E(S)$ and $\mathrm{Var}(S)$.

Solution 4.25 (a) The probability distribution of D, 'double the number on which the die lands', where $D = 2X$ is as shown:

d	2	4	6	8
$P(D = d)$	0.25	0.25	0.25	0.25

By symmetry, $E(D) = 5$.

$$\begin{aligned}\mathrm{Var}(D) &= E(D^2) - E^2(D)\\ &= \sum_{\text{all } d} d^2 P(D = d) - 25\\ &= 0.25(4 + 16 + 36 + 64) - 25\\ &= 5\end{aligned}$$

Therefore $E(D) = 5$ and $\mathrm{Var}(D) = 5$, where $D = 2X$.

NOTE:

$E(2X) = 5$	$\mathrm{Var}(2X) = 5$
$2E(X) = 2(2.5) = 5$	$4\mathrm{Var}(X) = 4(1.25) = 5$
So $E(2X) = 2E(X)$.	So $\mathrm{Var}(2X) = 4\mathrm{Var}(X)$.

(b) Consider the r.v. S where S is the sum of the two numbers on which the die lands when it is thrown twice. Therefore $S = X_1 + X_2$.

Now S can take the values 2, 3, 4, 5, 6, 7, 8 and the outcomes (all equally likely) are shown in the diagram:

Second throw				
4	5	6	7	8
3	4	5	6	7
2	3	4	5	6
1	2	3	4	5
	1	2	3	4
	First throw			

The probability distribution for S is:

s	2	3	4	5	6	7	8
$P(S = s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

By symmetry, $E(S) = 5$.

$$\begin{aligned}
\text{Var}(S) &= E(S^2) - E^2(S) \\
&= \sum_{\text{all } s} s^2 P(S = s) - 25 \\
&= \tfrac{1}{16}[4(1) + 9(2) + 16(3) + 25(4) + 36(3) + 49(2) + 64(1)] - 25 \\
&= 2.5
\end{aligned}$$

Therefore $E(S) = 5$, $\text{Var}(S) = 2.5$, where $S = X_1 + X_2$.

NOTE: $E(X_1 + X_2) = 5$

$2E(X) = 2(2.5) = 5$

So $E(X_1 + X_2) = 2E(X)$.

$\text{Var}(X_1 + X_2) = 2.5$

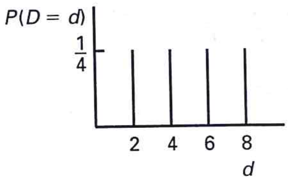
$2\text{Var}(X) = 2(1.25) = 2.5$

So $\text{Var}(X_1 + X_2) = 2\text{Var}(X)$.

We can see that the distribution for D, double the number on which the die lands, is very different from the distribution for S, the sum of the numbers on which the die lands when it is thrown twice.

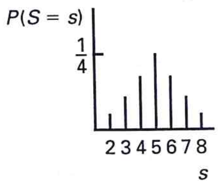
Double the number on one die:

d	2	4	6	8
$P(D = d)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$



The sum of the numbers when the die is thrown twice:

s	2	3	4	5	6	7	8
$P(S = s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$



Although the means of the two distributions are the same, the variances are not, with the r.v. 'double the number' having the greater variance.

Summarising, we have

MULTIPLES	SUMS
$E(2X) = 2E(X)$ $\text{Var}(2X) = 4\text{Var}(X)$	$E(X_1 + X_2) = 2E(X)$ $\text{Var}(X_1 + X_2) = 2\text{Var}(X)$
In general $E(nX) = nE(X)$ $\text{Var}(nX) = n^2\text{Var}(X)$	$E(X_1 + X_2 + \ldots + X_n) = nE(X)$ $\text{Var}(X_1 + X_2 + \ldots + X_n) = n\text{Var}(X)$

It is important that you understand whether multiples or sums are being considered. Think carefully about this point.

Exercise 4f

1. Independent random variables X and Y have probability distributions as shown in the tables:

x	0	1	2	3
$P(X = x)$	0.3	0.2	0.4	0.1

y	0	1	2
$P(Y = y)$	0.4	0.2	0.4

(*a*) Find $E(X)$, $E(Y)$, $\text{Var}(X)$, $\text{Var}(Y)$.
(*b*) Construct the probability distribution for the r.v. $X + Y$.
(*c*) Verify that $E(X + Y) = E(X) + E(Y)$.
(*d*) Verify that
$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
(*e*) Construct the probability distribution for the r.v. $X - Y$.
(*f*) Verify that $E(X - Y) = E(X) - E(Y)$.
(*g*) Verify that
$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.

2. Independent random variables X and Y are such that $E(X) = 4$, $E(Y) = 5$, $\text{Var}(X) = 1$, $\text{Var}(Y) = 2$. Find
(*a*) $E(4X + 2Y)$, (*b*) $E(5X - Y)$,
(*c*) $\text{Var}(3X + 2Y)$, (*d*) $\text{Var}(5Y - 3X)$,
(*e*) $\text{Var}(3X - 5Y)$.

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$X = 2$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

The above table gives the joint probability distribution of two random variables X and Y. Calculate (*a*) $P(Y = 1)$, (*b*) $P(XY = 2)$, (*c*) $E(X + Y)$. (L Additional)

4. Independent random variables X and Y are such that $E(X^2) = 14$, $E(Y^2) = 20$, $\text{Var}(X) = 10$, $\text{Var}(Y) = 11$. Find
(*a*) $E(3X - 2Y)$, (*b*) $\text{Var}(5X + 2Y)$.

5. Independent random variables X and Y are such that $E(X) = 3$, $E(X^2) = 12$, $E(Y) = 4$, $E(Y^2) = 18$. Find the value of
(*a*) $E(3X - 2Y)$, (*b*) $E(2Y - 3X)$,
(*c*) $E(6X + 4Y)$, (*d*) $\text{Var}(2X - Y)$,
(*e*) $\text{Var}(2X + Y)$, (*f*) $\text{Var}(3Y + 2X)$.

6. Two ordinary dice are thrown, a red and a green die. Let R be the r.v. 'the score on the red die' and let G be the r.v. 'the score on the green die'.
(*a*) Construct the probability distribution for $R + G$, the r.v. 'the sum of the two scores', and find (i) $E(R + G)$, (ii) $\text{Var}(R + G)$.
(*b*) Construct the probability distribution for $R - G$ and find (i) $E(R - G)$, (ii) $\text{Var}(R - G)$.
(*c*) Given that $E(R) = 3.5$ and $\text{Var}(R) = \frac{35}{12}$, comment on your answers.

7. X has probability distribution as shown:

x	0	1	2
$P(X = x)$	0.1	0.6	0.3

(*a*) Find $E(X)$ and $\text{Var}(X)$.
(*b*) Find $P(X_1 + X_2 = 4)$ where X_1, X_2 are two independent observations of X.
(*c*) Find $E(X_1 + X_2)$ and $\text{Var}(X_1 + X_2)$.
(*d*) Find $P(2X = 4)$.
(*e*) Find $E(2X)$ and $\text{Var}(2X)$.

8. Rods of length 2 m or 3 m are selected at random with probabilities 0.4 and 0.6 respectively.
(*a*) Find the expectation and variance of the length of a rod.
(*b*) Two lengths are now selected at random. Find the expectation and variance of the sum of the two lengths.
(*c*) Three lengths are now selected at random. Show that the probability distribution of Y, the sum of the three lengths, is:

y	6	7	8	9
$P(Y = y)$	0.064	0.288	0.432	0.216

and find $E(Y)$ and $\text{Var}(Y)$. Comment on your results.

9. Find the variance of the sum of the scores when an ordinary die is thrown 10 times.

10. X has a p.d.f. given by $P(X = x) = kx$, $x = 1, 2, 3, 4$. Find (*a*) k, (*b*) $E(X)$, (*c*) $\text{Var}(X)$, (*d*) $P(X_1 + X_2 = 5)$, (*e*) $E(4X)$, (*f*) $\text{Var}(X_1 + X_2 + X_3)$.

SUMMARY — DISCRETE RANDOM VARIABLES

For the discrete random variable X with probability density function $P(X=x)$ for $x = x_1, x_2, \ldots, x_n$,

(1)	$\sum_{\text{all } x} P(X=x) = 1$
(2)	$F(t) = \sum_{x=x_1}^{t} P(X=x)$ where $F(t)$ is the cumulative distribution function
(3)	$\mu = E(X) = \sum_{\text{all } x} xP(X=x)$ where μ is the expectation of X
(4)	$\text{Var}(X) = E(X^2) - E^2(X) = \sum_{\text{all } x} x^2P(X=x) - \mu^2$

For the random variable X and constants a and b,

$E(a) = a$	$\text{Var}(a) = 0$
$E(aX) = aE(X)$	$\text{Var}(aX) = a^2\text{Var}(X)$
$E(aX+b) = aE(X)+b$	$\text{Var}(aX+b) = a^2\text{Var}(X)$

For any two random variables X and Y and constants a and b,

$E(X+Y) = E(X)+E(Y)$

$E(X-Y) = E(X)-E(Y)$

$E(aX+bY) = aE(X)+bE(Y)$

$E(aX-bY) = aE(X)-bE(Y)$

For independent random variables X and Y and constants a and b,

$\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$

$\text{Var}(X-Y) = \text{Var}(X)+\text{Var}(Y)$

$\text{Var}(aX+bY) = a^2\text{Var}(X)+b^2\text{Var}(Y)$

$\text{Var}(aX-bY) = a^2\text{Var}(X)+b^2\text{Var}(Y)$

If $X_1, X_2, \ldots, X_n$ are n independent observations of the r.v. X then,

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

$\text{Var}(X_1+X_2+\ldots+X_n) = n\text{Var}(X)$

Miscellaneous Exercise 4g

1. Two tetrahedral dice are thrown and the score is the product of the numbers on which the dice fall. What is the expected score for a throw?

2. A woman removes the labels from three tins of peaches and a tin of baked beans in order to enter a competition and then puts the tins in a cupboard. She discovers that the tins are outwardly identical. Let X be the number of tins she now needs to open in order to have baked beans. List the values that X can take and determine the probabilities for each of these values of X. Calculate the expected value of X.
Her neighbour has five tins of peaches and two tins of baked beans, again outwardly identical once the labels are removed. This woman removes the labels and puts the tins away. Find the probability that this woman later requires to open at least three tins to have baked beans.

3. On a long train journey, a statistician is invited by a gambler to play a dice game. The game uses two ordinary dice which the statistician is to throw. If the total score is 12, the statistician is paid £6 by the gambler. If the total score is 8, the statistician is paid £3 by the gambler. However if both or either dice show a 1, the statistician pays the gambler £2. Let £X be the amount paid to the statistician by the gambler after the dice are thrown once.
Determine the probability that (*a*) $X = 6$, (*b*) $X = 3$, (*c*) $X = -2$.
Find the expected value of X and show that, if the statistician played the game 100 times, his expected loss would be £2.78, to the nearest penny.
Find the amount, £a, that the £6 would have to be changed to in order to make the game unbiased.

4. A box contains nine numbered balls. Three balls are numbered 3, four balls are numbered 4 and two balls are numbered 5. Each trial of an experiment consists of drawing two balls without replacement and recording the sum of the numbers on them, which is denoted by X. Show that the probability that $X = 10$ is $\frac{1}{36}$, and find the probabilities of all other possible values of X. Use your results to show that the mean of X is $\frac{70}{9}$, and find the standard deviation of X. Two trials are made. (The two balls in the first trial are replaced in the box before the second trial.) Find the probability that the second value of X is greater than or equal to the first value of X. (MEI)

5. A man stakes £2 to play a game in which he rolls an ordinary (fair) die. If he scores 1 or 2 he wins £3 (plus his stake) and loses his stake if he scores 3, 4 or 5. If he scores a six he may roll the die once again, winning if he scores 1, 2 or 6, losing if he scores 3, 4 or 5.
Find (*a*) the probability that the man wins the game by rolling (i) once, (ii) twice;
(*b*) his expected gain,
(*c*) the expected number of times he will roll the die.
If the rules are changed so that the winning scores are 1 and 2 but that every time he scores 6 he may roll the die again, find
(*d*) the probability that he wins on his rth roll of the die,
(*e*) the probability that he wins the game.

6. A and B each roll a fair die simultaneously. Construct a table for the difference in their scores showing the associated probabilities. Calculate the mean of the distribution. If the difference in scores is 1 or 2, A wins; if it is 3, 4 or 5, B wins and if it is zero, they roll their dice again. The game ends when one of the players has won. Calculate the probability that A wins on (*a*) the first, (*b*) the second, (*c*) the rth roll. What is the probability that A wins?
If B stakes £1 what should A stake for the game to be fair?

7. A gambler has 4 packs of cards, each of which is well shuffled and has equal numbers of red, green and blue cards. For each turn he pays £2 and draws a card from each pack. He wins £3 if he gets 2 red cards, £5 if he gets 3 red cards and £10 if he gets 4 red cards.
(*a*) What are the probabilities of his drawing 0, 1, 2, 3, 4 red cards?
(*b*) What is the expectation of his winnings (to the nearest 10 p)?

8. During winter a family requests 4 bottles of milk every day, and these are left on the door-step. Three of the bottles have silver tops and the fourth has a gold top. A thirsty blue-tit attempts to remove the tops from these bottles. The probability distribution of X, the number of silver tops removed by the blue-tit, is the same each day and is given by

$$P(X=0) = \frac{5}{15}, \quad P(X=1) = \frac{6}{15}, \quad P(X=2) = \frac{3}{15}, \quad P(X=3) = \frac{1}{15}$$

The blue-tit finds the gold top particularly attractive, and the probability that this top is removed is $\frac{3}{5}$, independent of the number of silver tops removed. Determine the expectation and variance of

(a) the number of silver tops removed in a day,

(b) the number of gold tops removed in a day,

(c) the total number of tops (silver and gold) removed in 7 days.

Find also the probability distribution of the total number of tops (silver and gold) removed in a day. (C)

9. The probability of there being X unusable matches in a full box of Surelite matches is given by $P(X=0) = 8k$, $P(X=1) = 5k$, $P(X=2) = P(X=3) = k$, $P(X \geqslant 4) = 0$. Determine the constant k and the expectation and variance of X.
Two full boxes of Surelite matches are chosen at random and the total number Y of unusable matches is determined. Calculate $P(Y > 4)$, and state the values of the expectation and variance of Y. (C)

10. A player throws a die whose faces are numbered 1 to 6 inclusive. If the player obtains a six he throws the die a second time, and in this case his score is the sum of 6 and the second number; otherwise his score is the number obtained. The player has no more than two throws.
Let X be the random variable denoting the player's score. Write down the probability distribution of X, and determine the mean of X.
Show that the probability that the sum of two successive scores is 8 or more is $\frac{17}{36}$.
Determine the probability that the first of two successive scores is 7 or more, given that their sum is 8 or more. (C)

11. The faces of an ordinary die are renumbered so that the faces are 1, 2, 2, 3, 3 and 3. This die and an ordinary, unaltered die are thrown at the same time. The score, X, is the sum of the numbers on the uppermost faces of the two dice. Show that the probability of X being 3 is $\frac{1}{12}$ and of being 4 is $\frac{1}{6}$.
List the values that X can take and determine their respective probabilities.
Hence obtain the expected value of X, correct to 3 decimal places.
If the dice are thrown 3 times, determine the probability, correct to 3 significant figures, that none of the three values of X exceeds 3.

12. Alan and his younger brother Bill play a game each day. Alan throws three darts at a dartboard and for each dart that scores a bull (which happens with probability p) Bill gives him a penny, while for each dart which misses the bull (which happens with probability $1-p$) Alan gives Bill two pence. By considering all possible outcomes for the three throws, or otherwise, find the distribution of the number of pence (positive or negative) that Bill receives each day. Show that, when $p = \frac{1}{3}$, the mean is 3 and the variance 6.
The game takes place on 150 days. What is the mean and standard deviation of Bill's total winnings when $p = \frac{1}{3}$? (O)

13. In a certain field, each puffball which is growing in one year gives rise to a number, X, of new puffballs in the following year. None of the original puffballs is present in the following year. The probability distribution of the random variable X is as follows:

$$P(X=0) = P(X=2) = 0.3,$$

$$P(X=1) = 0.4.$$

Find the probability distribution of Y, the number of puffballs resulting from there being two puffballs in the previous year, and show that the variance of Y is 1.2.
Hence, or otherwise, determine the probability distribution of the number, Z, of puffballs present in year 3, given that there was a single puffball present in year 1. Find also the mean and variance of Z. (C)

14. A discrete random variable X can take only the values 0, 1, 2 or 3, and its probability distribution is given by $P(X=0) = k$, $P(X=1) = 3k$, $P(X=2) = 4k$, $P(X=3) = 5k$, where k is a constant. Find

(a) the value of k,

(b) the mean and variance of X. (JMB)

15. A random variable R takes the integer value r with probability $P(r)$ where

$$P(r) = kr^3, \qquad r = 1, 2, 3, 4,$$

$$P(r) = 0, \qquad \text{otherwise}$$

Find

(a) the value of k, and display the distribution on graph paper,

(b) the mean and the variance of the distribution,

(c) the mean and the variance of $5R-3$. (L)P

16. A gambling machine works in the following way. The player inserts a penny into one of five slots, which are coloured Blue, Red, Orange, Yellow and Green corresponding to five coloured light bulbs. The player can choose whichever coloured slot he likes. After the penny has been inserted one of the five bulbs lights up. If the bulb lit up is the same colour as the slot selected by the player, then the player wins and receives from the machine R pennies, where

$P(R = 2) = \frac{1}{2}$, $P(R = 4) = \frac{1}{4}$

$P(R = 6) = \frac{3}{20}$, and

$P(R = 8) = P(R = 10) = \frac{1}{20}$

If the colour of the bulb lit up and the slot selected are not the same, the player receives nothing from the machine. In either case the player does not get back the penny that he inserted. Assuming that each of the colours is equally likely to light up, and that the machine selects the bulbs at random, determine

(*a*) the probability that the player receives nothing from the machine,

(*b*) the expected value of the amount gained by the player from a single try,

(*c*) the variance of the amount gained by the player from a single try. (C)

17. Four rods of lengths 1, 2, 3 and 4 units are placed in a bag from which one rod is selected at random. The probability of selecting a rod of length l is kl. Find the value of k.
Show that the expected value of X, the length of the selected rod, is 3 units and find the variance of X.
After a rod has been selected it is not replaced. The probabilities of selection for each of the three rods that remain are in the same ratio as they were before the first selection. A second rod is now selected from the bag. Defining Y to be the length of this rod and writing $P_1 = P(Y = 1|X = 2)$, $P_2 = P(Y = 2|X = 1)$ show that $16P_1 = 9P_2$.
Show also that $P(X + Y = 3) = \frac{17}{360}$. (C)

18. A game is played in which a complete throw consists of three fair coins being tossed once each and any which have landed tails being tossed a second time; no coin is tossed more than twice. The score for the complete throw is the total number of heads showing at the end of the throw.
(*a*) Find the respective probabilities that the score after a complete throw is
(i) 0, (ii) 1, (iii) 2, (iv) 3.
(*b*) Show that the average score over a large number of complete throws is $\frac{9}{4}$.
(You may leave your answers as fractions in their lowest terms.) (O & C)

19. The random variable X takes values $-2, 0, 2$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ respectively. Find $\mathrm{Var}(X)$ and $E(|X|)$.
The random variable Y is defined by $Y = X_1 + X_2$, where X_1 and X_2 are two independent observations of X. Find the probability distribution of Y. Find $\mathrm{Var}(Y)$ and $E(Y + 3)$. (C)

20. The discrete random variable X can take only the values 0, 1, 2, 3, 4, 5. The probability distribution of X is given by the following:

$P(X = 0) = P(X = 1) = P(X = 2) = a$

$P(X = 3) = P(X = 4) = P(X = 5) = b$

$P(X \geqslant 2) = 3P(X < 2)$

where a and b are constants.
(i) Determine the values of a and b.
(ii) Show that the expectation of X is $\frac{23}{8}$ and determine the variance of X.
(iii) Determine the probability that the sum of two independent observations from this distribution exceeds 7. (C)

21. A random variable R takes the integer values $1, 2, \ldots, n$ each with probability $1/n$. Find the mean and variance of R.
A pack of 15 cards bearing the numbers 1 to 15 is shuffled. Find the probability that the number on the top card is larger than that on the bottom card, giving reasons for your answer.
If the sum of these two numbers is S, find
(*a*) the probability that $S \leqslant 4$,
(*b*) the expected value of S.
(Answers may be left as fractions in their lowest terms.) (O & C)

22. A discrete random variable X has the distribution function

x	1	2	4	5
$F(x)$	$\frac{1}{12}$	$\frac{1}{2}$	$\frac{5}{6}$	1

(*a*) Write down the probability distribution of X.
(*b*) Find the probability distribution of the sum of two independent observations from X and find the mean and variance of the distribution of this sum.

23. A random variable R takes the integer value r with probability $P(r)$ defined by

$P(r) = kr^2$, $r = 1, 2, 3,$

$P(r) = k(7 - r)^2$, $r = 4, 5, 6,$

$P(r) = 0$, otherwise.

Find the value of k and the mean and variance of the probability distribution.
Exhibit this distribution by a suitable diagram.
Determine the mean and the variance of the variable Y where $Y \equiv 4R - 2$. (L)P

24. A curiously shaped six-faced die produces a score, X, for which the probability distribution is given in the following table.

r	1	2	3	4	5	6
$P(X = r)$	k	$k/2$	$k/3$	$k/4$	$k/5$	$k/6$

Show that the constant k is 20/49. Find the mean and variance of X.
The die is thrown twice. Show that the probability of obtaining equal scores is approximately $\frac{1}{4}$. (MEI)

25. A random number generator in a computer game produces values which can be modelled by the discrete random variable X with probability distribution given by

$$P(X = r) = kr! \qquad r = 0, 1, 2, 3, 4,$$

where k is a constant.
(i) Show that $k = 1/34$ and illustrate the probability distribution with a sketch.
(ii) Find the expectation and variance of X.
Two independent values of X are generated. Let these values be X_1 and X_2.
(iii) Show that $P(X_1 = X_2)$ is a little greater than 0.5.
(iv) Given that $X_1 = X_2$, find the probability that X_1 and X_2 are each equal to 4. (MEI)

26. Write down an expression involving probabilities for $P(B \mid A)$, the probability of event B given that event A occurs.
Alison and Brenda play a tennis match in which the first player to win two sets wins the match. In tennis no set can be drawn. The probability that Alison wins the first set is $\frac{1}{3}$; for sets after the first, the probability that Alison wins the set is $\frac{3}{5}$ if she won the preceding set, but is only $\frac{1}{4}$ if she lost the preceding set.
With the aid of a suitable diagram, or otherwise, determine the probability that
(i) the match lasts for just two sets,
(ii) Alison wins the match given that it lasts for just two sets,
(iii) Alison wins the match,
(iv) Alison wins the match given that it goes to three sets,
(v) if Alison wins the match, then she does so in two sets.
Calculate the expected number of sets that Alison wins. (JMB)

27. (*a*) A regular customer at a small clothes shop observes that the number of customers, X, in the shop when she enters has the following probability distribution.

Number of customers, x	0	1	2	3	4
Probability $p(x)$	0.15	0.34	0.27	0.14	0.10

(i) Find the mean and standard deviation of X.
She also observes that the average waiting time, Y, before being served, is as follows.

Number of customers, x	0	1	2	3	4
Average waiting time, v minutes	0	2	6	9	12

(ii) Find her mean waiting time.
(*b*) The customer decides that, in future, if there are more than two customers waiting when she arrives she will leave immediately and return another day. However when she returns she will wait no matter how long the queue is.
(i) What is the probability of her leaving immediately on her first visit?
(ii) Given that she leaves immediately, what is the probability of there being more customers in the shop when she returns than on her first visit?
(iii) Find her mean waiting time under these conditions.
(*c*) The customer now decides that, if there are more than 2 people in the shop she will not wait but will return on another day no matter how many times this occurs. What is the probability that she will have to make more than 3 visits before being served? (AEB 1989)