4 DISCRETE PROBABILITY DISTRIBUTIONS

Objectives

After studying this chapter you should

- understand what is meant by a discrete probability distribution;
- be able to find the mean and variance of a distribution;
- be able to use the uniform distribution.

4.0 Introduction

The definition

'$X =$ the total when two standard dice are rolled'

is an example of a random variable, X, which may assume any of the values in the range 2, 3, 4, ..., 12. The outcome cannot be predicted with certainty though probabilities can be assigned to each possible result.

A **random variabl**e is a quantity that may take any of a given range of values that cannot be predicted exactly but can be described in terms of their probability. As was seen in Chapter 2, data is classified either as **discrete** if the values are taken from a fixed number of numerical values (generally assessed by counting), or **continuous** if the values can fall anywhere over a range and the scale is only restricted by the accuracy of measuring. Some examples of data which can be described by a random variable are shown below.

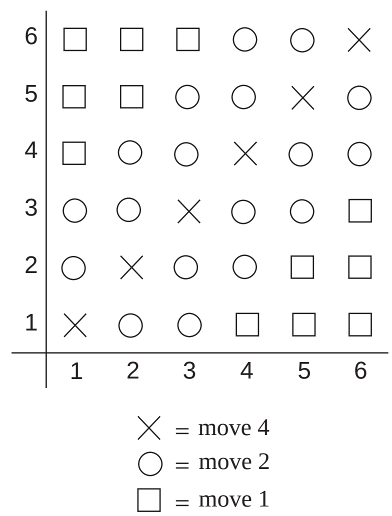
Discrete	Continuous
number of red smarties in a packet	weight of babies at birth
number of traffic accidents in Leeds in one day	lengths of pine cones in a wood
number of throws required to score 6 with a single die	time needed to drive from Lincoln to Dover

Discuss whether the times taken to run 100 m in the Olympics will be values of a discrete rather than a continuous random variable.

4.1 Expectation

Activity 1

Play a game in which a counter is moved forward one, two or four places according to whether the scores on the two dice rolled differ by three or more, by one or two, or are equal. Here is a random variable, M, the number of places moved, which can take the values 1, 2 or 4. Play the game at least 20 times and evaluate from the games the average (mean) number of moves per game.



The probabilities of each of these values occurring can be calculated from the diagram opposite. Check for yourself that the probabilities are in fact,

$$P(M=1) = \frac{12}{36} = \frac{1}{3}$$

$$P(M=2) = \frac{18}{36} = \frac{1}{2}$$

$$P(M=4) = \frac{6}{36} = \frac{1}{6}.$$

In the long run you would expect to move one square $\frac{1}{3}$ of the times, two squares on $\frac{1}{2}$ of the goes and four squares on $\frac{1}{6}$. So if you play the game 36 times you will expect to average a total of

$$\begin{aligned} 1\times 12 + 2\times 18 + 4\times 6 &= 12 + 36 + 24 \\ &= 72 \text{ moves.} \end{aligned}$$

If you divide by 36 to get the mean number of moves per game the equation becomes

$$\frac{1\times 12 + 2\times 18 + 4\times 6}{36} = \frac{72}{36}$$

which you can write as

$$1\times\frac{12}{36} + 2\times\frac{18}{36} + 4\times\frac{6}{36} = 2.$$

So $\quad$ mean $= 1\times P(M=1) + 2\times P(M=2) + 4\times P(M=4)$

$\qquad\qquad = 2.$

Using the summation symbol , $\sum$, the last equation can then be shortened to

$$\text{mean} = \sum_{\text{all } m} m \times P(M = m)$$

where $m = 1, 2, 4$ are the possible values taken by *M*.

The quantity 2 is the **mean** or **expectation** or **expected value** of the random variable *M*, written $E(M)$, in the example above.

In general, for a discrete random variable *X*, which can take specific values of *x*, the expected value (mean) of the random variable is defined by

$$\boxed{E(X) = \sum_{\text{all } x} x \times P(X = x)}$$

where the summation over 'all *x*' means all values of *x* for which the random variable *X* has a non-zero probability.

Example

When throwing a normal die, let *X* be the random variable defined by

$$X = \text{ the square of the score shown on the die.}$$

What is the expectation of *X*?

Solution

The possible values of *X* are

$$1,\quad 2^2,\quad 3^2,\quad 4^2,\quad 5^2 \text{ and } 6^2$$
$$\Rightarrow \quad 1,\quad 4,\quad 9,\quad 16,\quad 25 \text{ and } 36.$$

Each one has a probability of $\frac{1}{6}$ of occurring, so

$$E(X) = 1\times\frac{1}{6} + 4\times\frac{1}{6} + 9\times\frac{1}{6} + 16\times\frac{1}{6} + 25\times\frac{1}{6} + 36\times\frac{1}{6}$$

$$= \frac{1}{6}\times 91$$

$$= 15\frac{1}{6}.$$

Finally in this section, an alternative definition of a random variable will be developed.

In the previous example, what is the value of $\sum_{\text{all } x} P(X = x)$?

If the summation is over all possible values of x, the summation must add up to one. So an alternative way of defining a discrete random variable is to impose the condition

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

Exercise 4A

1. Categorise each of the following as continuous or discrete. The random variables are:

   (a) A is 'the age in completed years of the first person I see wearing a hat'.

   (b) B is 'the length of the next car to enter the car park'.

   (c) C is 'how many cows I will see before the first green one'.

   (d) D is 'the date next July of the day with the highest temperature'.

2. Let $X =$ total score when two dice are rolled.

   (a) Find the possible values of the random variable X and determine the associated possibilities.

   (b) Determine the expectation of X.

   Could you answer (b) without actually performing any calculations?

3. Let $X =$ number of heads obtained when tossing a fair coin 3 times.

   (a) What are the possible values of X?

   (b) What are the associated probabilities?

   (c) Determine the mean value of X.

4. The random variable X has the probability distribution shown.

x	1	2	3	4	5
$P(X = x)$	$7c$	$5c$	$4c$	$3c$	c

   (a) Use the fact that $\sum P(X = x) = 1$ to find c.

   (b) Explain why you expect $E(X)$ to be 3, greater than 3, or less than 3.

   (c) Calculate $E(X)$.

5. The random variable Z has probability distribution:

z	2	3	5	7	11
$P(Z = z)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{4}$	x	y

   and $E(Z) = 4\frac{2}{3}$. Find x and y.

4.2 Variance

The expression for the expected value just produced looks very similar to that in Chapter 3 which gave

$$\bar{x} = \frac{\sum x}{n}$$

for the mean value of a set of data.

A formula for variance like that from Chapter 3 can also be derived. Note that the variance was defined by

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2.$$

In the example in the previous section, groups of 36 terms could be expected, on average, to give a variance of

$$\frac{1^2 \times 12 + 2^2 \times 18 + 4^2 \times 6}{36} - 2^2$$

$$= 1^2 \times \frac{12}{36} + 2^2 \times \frac{18}{36} + 4^2 \times \frac{6}{36} - 2^2$$

$$= 1^2 P(M=1) + 2^2 P(M=2) + 4^2 P(M=4) - 2^2.$$

In general, the variance is defined by

$$\boxed{V(X) = E\left(X^2\right) - \left[E(X)\right]^2}$$

and the standard deviation, σ, is as before defined by

$$\sigma^2 = V(X).$$

For the example above,

$$V(M) = E\left(M^2\right) - \left[E(M)\right]^2$$

$$= \sum m^2 \, P(M=m) - \left[\sum m \, P(M=m)\right]^2,$$

giving $$V(M) = 1^2 \times \frac{1}{3} + 2^2 \times \frac{1}{2} + 4^2 \times \frac{1}{6} - 2^2$$

$$= \frac{1}{3} + 2 + \frac{8}{3} - 4$$

$$= 1.$$

As with data, the standard deviation gives a measure of the spread of the distribution.

Example

Find the variance and standard deviation of X, where

X = the square of the score shown on a die.

Solution

The possible values of X are

$$1,\ 4,\ 9,\ 16,\ 25 \text{ and } 36$$

each one having a probability of $\frac{1}{6}$. As you saw in Section 4.1, the mean value (expectation) is given by

$$E(X) = \tfrac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = 15\tfrac{1}{6}$$

whilst

$$E\left(X^2\right) = \tfrac{1}{6}\times 1 + \tfrac{1}{6}\times 16 + \tfrac{1}{6}\times 81 + \tfrac{1}{6}\times 256 + \tfrac{1}{6}\times 625 + \tfrac{1}{6}\times 1296$$

$$= \tfrac{1}{6}\times 2275$$

$$= 379\tfrac{1}{6}.$$

Thus $$V(X) = \frac{2275}{6} - \left(\frac{91}{6}\right)^2 = \frac{13650 - 8281}{36}$$

$$= \frac{5369}{36}$$

and $$\sigma = \sqrt{\frac{5369}{36}} \approx 12.2.$$

Exercise 4B

1. If X is the score on a fair die, find the variance and standard deviation of X.
2. Find the variance and standard deviation of X when X is defined as in:

   (a) Exercise 4A, Question 2;

   (b) Exercise 4A, Question 3.
3. A team of 3 is to be chosen from 3 boys and 4 girls. If X is the random variable 'the number of girls in the team', find its probability distribution and hence find:

   (a) $E(X)$;

   (b) $V(X)$ and the standard deviation.

4.3 Probability distributions

The probabilities for the game in Activity 1 can be written in a table as:

m	1	2	4
$P(M=m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

This gives the **probability distribution** of M as it shows how the total probability of 1 is distributed over the possible values. The probability distribution is often denoted by $p(m)$.

So $\quad p(1) = P(M=1) = \dfrac{1}{3}, \quad p(2) = \dfrac{1}{2}, \quad p(3) = \dfrac{1}{6}.$

In general, $P(X=x) = p(x)$, and p can often be written as a formula.

Example

The discrete random variable X has probability distribution

$$p(x) = \frac{x}{36} \quad \text{for } x = 1, 2, 3, \ldots, 8.$$

Find $E(X)$ and $V(X)$.

Solution

Substituting the values 1 to 8 into the probability distribution gives

x	1	2	3	4	5	6	7	8
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{7}{36}$	$\frac{8}{36}$

(The probability distribution is a shorter way of giving all the probabilities associated with the random variable than drawing up a table, and indeed, there is no need to write one out if you do not feel it helps.)

As expected, note that the sum of all the probabilities is 1.

$$\begin{aligned} E(X) &= \sum x\, p(x) \\ &= 1 \times \frac{1}{36} + 2 \times \frac{2}{36} + 3 \times \frac{3}{36} + \ldots + 8 \times \frac{8}{36} \\ &= \frac{(1 + 4 + 9 + \ldots + 64)}{36} \\ &= \frac{204}{36} \\ &= 5\frac{2}{3}. \end{aligned}$$

Does this seem likely?

Well, the values five to eight have greater probabilities than one to four so the expected answer should be more than $4\frac{1}{2}$.

$$V(X) = \left(1^2 \times \frac{1}{36} + 2^2 \times \frac{2}{36} + \ldots + 8^2 \times \frac{8}{36}\right) - \left(5\frac{2}{3}\right)^2$$

$$= \frac{(1 + 8 + 27 + \ldots + 512)}{36} - 32\frac{1}{9}$$

$$= 36 - 32\frac{1}{9}$$

$$= 3\frac{8}{9}.$$

In the following section, you will consider some special probability distributions which have wide applicability.

Exercise 4C

1. For a discrete random variable Y the probability distribution is

$$p(y) = \frac{5-y}{10} \quad \text{for} \quad y = 1,\ 2,\ 3,\ 4.$$

Calculate: (a) $E(Y)$ (b) $V(Y)$.

2. For a fair 10-sided spinner, if S is 'the score on the spinner', find:

(a) the probability distribution of S; (b) $E(S)$;

(c) the standard deviation of S.

3. A random variable has probability distribution

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

Find:

(a) the mean and variance of X;

(b) the mean and variance of the random variable

$$Y = X^2 - 2X.$$

4. A fair six-sided die has

'1' on one face
'2' on two of its faces
'3' on the remaining three faces.

The die is thrown twice, and X is the random variable 'total score thrown'. Find

(a) the probability distribution;

(b) the probability that the total score is more than 4;

(c) $E(X)$ and $V(X)$.

4.4 The uniform distribution

One important distribution is the uniform one in which all possible outcomes have equal possibilities.

Activity 2 A survey of car registration plates

Survey vehicles in a car park or at any convenient place and note the digits on the number plates. Draw up a table like the one shown opposite.

What do you notice about the distribution of digits?

Digit	Tally	Frequency	Relative frequency
0			
1			
2			
.			
.			
.			
9			

The random variable X is said to follow a **uniform distribution** when all its outcomes are equally likely. A very simple example is given by the random variable H, 'the number of heads seen when a single coin is tossed'.

The probability distribution is given by

h	0	1
$p(h)$	$\frac{1}{2}$	$\frac{1}{2}$

or $$p(h) = \frac{1}{2} \quad \text{for } h = 0, 1.$$

So $$E(H) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$$

and $$V(H) = 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} - \left(\frac{1}{2}\right)^2$$
$$= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

and the standard deviation of H is $\frac{1}{2}$.

Activity 3 How random is your calculator?

Computers and certain calculators have a facility to enable you to generate random numbers. Some calculators will produce random numbers to three decimal places from 0.000 to 0.999. By simply reading only the first figure after the decimal point you can produce a set of random digits which should be uniformly distributed. Use your calculator to produce one hundred random digits and draw a frequency diagram to see how close your results are to the expected values.

Another example of a uniform distribution is the random variable, X, the score obtained when rolling a single unbiased die. In this case

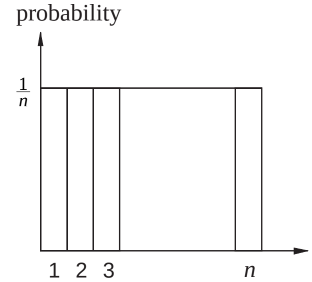
$$P(X=x) = \frac{1}{6} \quad \text{for} \quad x = 1, 2 \ldots, 6$$

giving $$E(X) = \frac{7}{2} = 3\frac{1}{2},$$

which can be written down by considering symmetry, and

$$V(X) = \frac{1}{6}(1+4+9+16+25+36) - \frac{49}{4}$$

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



This can be generalised to a random variable, X, having n equally likely outcomes for which the probability distribution is given by

$$\boxed{P(X=x) = \frac{1}{n} \quad \text{for} \quad x = 1, 2 \ldots, n}$$

Activity 4

For a 30-sided spinner, let X be the score obtained. Determine

(a) $E(X)$ (b) $V(X)$.

You may be able to spot patterns in the results with n equal to 6 and 30 and deduce the general results.

i.e. $n = 6$ $\quad E(X) = \frac{7}{2}$ $\quad V(X) = \frac{35}{12}$

$n = 30$ $\quad E(X) = \frac{31}{2}$ $\quad V(X) = \frac{899}{12}$

n $\quad E(X) = ?$ $\quad V(X) = ?$

The number 899 looks 'clumsy' but it can be written as

$900-1=30^2-1$, which suggests $V(X)=\dfrac{n^2-1}{12}$, although $E(X)=\dfrac{n+1}{2}$ is simpler to see.

You will need the series summation results:

$$1+2+3+\ldots+n = \frac{n}{2}(n+1) \qquad (1)$$

and

$$1^2+2^2+3^2+\ldots+n^2 = \frac{n}{6}(n+1)(2n+1) \qquad (2)$$

to understand the proofs of the general results.

So
$$E(X) = 1\times\frac{1}{n} + 2\times\frac{1}{n} + 3\times\frac{1}{n} + \ldots + n\times\frac{1}{n}$$

$$= (1+2+3+\ldots+n)\times\frac{1}{n}, \quad \text{using (1)},$$

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

$$\boxed{E(X) = \frac{n+1}{2}}$$

Activity 5

Show, using equation (2), that

$$V(X) = \frac{n^2-1}{12}.$$

4.5 Miscellaneous Exercises

1. Find the probability distribution for the random variable
   (a) the number of sixes obtained when two ordinary dice are thrown,
   (b) the smaller or equal number when two ordinary dice are thrown,
   (c) the number of heads when three fair coins are tossed.

2. For the discrete random variable X, the probability distribution is given by

$$P(X = x) = \begin{cases} kx & x = 1,2,3,4,5 \\ k(10 - x) & x = 6,7,8,9 \end{cases}$$

Find:

(a) the value of the constant k

(b) $E(x)$

(c) $V(x)$

3. 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 possible 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.

4. 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. (SUJB)

5. 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? (SUJB)

6. 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 10p)? (SUJB)

7. A player throws a die whose faces are numbered 1 to 6 inclusively. 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 not 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.

8. 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:

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

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

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

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.