7 CONTINUOUS PROBABILITY DISTRIBUTIONS

Objectives

After studying this chapter you should

• understand the use of continuous probability distributions and the use of area to calculate probabilities;

• be able to use probability functions to calculate probabilities and find measures such as the mean and variance;

• recognise and be able to use the rectangular distribution.

7.0   Introduction

Note that in order to work through this chapter you will need to be able to

(a)   factorise and expand polynomials up to order 3;

(b)   integrate simple functions and use definite integrals to find areas under curves;

(c)   differentiate simple functions and find turning points.

On virtually every food item purchased you will find a **nominal** weight. If you find a packet of crisps which weighs 24 g and the nominal weight is 25 g, are you entitled to complain? Clearly manufacturers cannot be expected to make every packet exactly 25 g but the law requires a certain percentage of all packets to be above this weight. The manufacturer therefore needs to know the pattern or distribution of the weights of the crisp packets in order to check whether or not the company is breaking the law.

Activity 1   Coin tossing

For this you will need a number of 2p coins. Mark out a playing grid with pieces of string on a tarmac area or on short grass, as shown opposite.

The central target line could be drawn in a different colour. The aim of the game is for a person standing on the base line to toss the coins to land as close to the target line as possible. Any coins falling outside the grid area are taken again. Let each member of
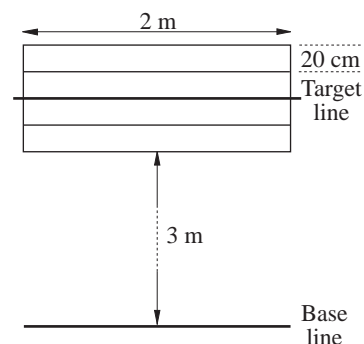
the group try it a number of times to give about 100 results. Record the distance each coin lands away from the target line in centimetres noting whether it is in front or behind with $-/+$ respectively.

If you wanted to write a computer program to 'simulate' this game you would need to know the probability of the coin landing different distances from the line. Provided your aim is fairly good you should expect to get more shots nearer the line than further away. The same idea is used in evaluating how likely an artillery weapon is to hit its target. Gunners can only estimate the distance to a target and shots will fall in a particular pattern around the target.

The main aim of this chapter is to develop a method of representing the probabilities in terms of a continuous function. This will enable estimates to be made as to what proportion will be within specified limits.

7.1   Looking at the data

In a similar experiment shot-putters were asked to aim at a line 10 m away. They threw the shot 200 times and throws were measured within 2 m either side of the line. The results were as shown below.

	in front			
metres	1.99-1.50	1.49-1.00	0.99-0.5	0.49-0
frequency	9	22	31	37

	behind			
metres	0-0.49	0.50-0.99	1.00-1.49	1.50-1.99
frequency	38	32	23	8

A histogram to represent these data is shown opposite.

You could say that this pattern was the one that all throws were likely to follow. To obtain a more accurate picture it might be possible to collect more data; this would also allow narrower groups to be used. A possible pattern which might emerge is shown on the next page.
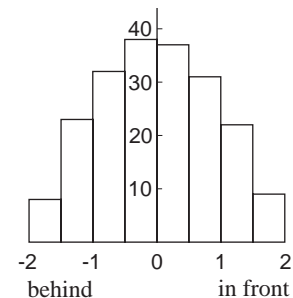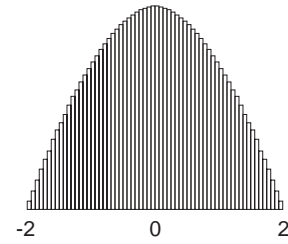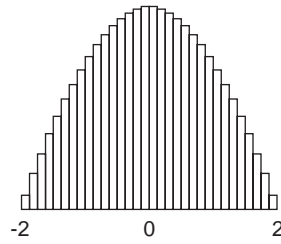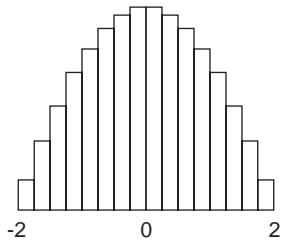
In the final case the bars are so thin it appears that the tops are a continuous curve - this is called a **frequency curve**. For very large samples then, the graph can be shown as opposite.

Unless you know the total sample size though, you cannot put a scale on the *y*-axis. So what should you use?
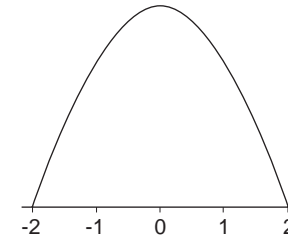
In how many throws out of 100 roughly would a shot putter throw a distance of exactly 10 m?

How many throws in 100 would you expect to be between 9.9 m and 10.1 m?

How many throws in 100 would be behind the target line?

How many would be between 1 m and 2 m beyond the line?

How would the answer differ for 500 throws or 5000?

Two factors should emerge from this, namely that

(a)   the proportion of throws getting a precise value is infinitely small;

(b)   for a particular range of values the **proportion** remains constant.

In order to answer some of the above questions more accurately you will need to measure the areas under the graph. From the original data the probability of any throw being between 0 and 1 m behind the line is
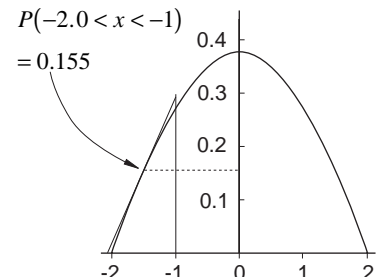
$$\frac{(38+32)}{200} = 0.35$$
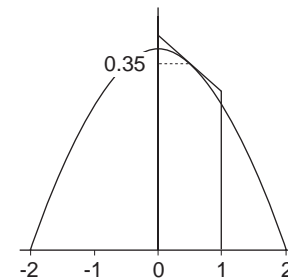
On the graph we can estimate this by using a trapezium, which will have mid-point height 0.35 to give an area 0.35.

Using all the other results a scale can be given to the vertical axis. Note that, as shown opposite,
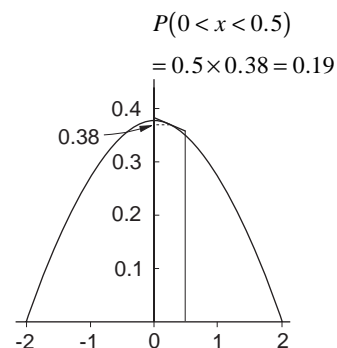
$$P(-2.0 < x < -1.0) = 0.155.$$

$P(-2.0 < x < -1)$
$= 0.155$

For smaller ranges the area principle still works; for example

$$P(0 < x < 0.5) = 0.5 \times 0.38 = 0.19.$$

Such graphs as these are called **probability distributions** and they can be used to find the probability of a particular **range** of values occurring.

$P(0 < x < 0.5)$

$= 0.5 \times 0.38 = 0.19$

Activity 2   Putting your results together

Draw a histogram of your coin tossing results using 10 cm intervals. Draw a frequency curve of your results and scale with relative probabilities. Remember that for a 10 cm group the percentage frequencies will have to be divided by 10. Use your graph and trapeziums to find approximately:

(a)   the probability of a coin landing between 10 and 30 cm past the line;

(b)   the probability of a throw being more than 20 cm past the line;

(c)   the probability of a throw being more than 15 cm short of the line.

7.2   Finding a function

Activity 3   Fishing for a function

The data on the page opposite represent the fish caught in one trawl by a scientific research vessel. These are all shown at 1/10th scale. A scientist wants to set up a probability distribution for the lengths of fishes so that she can simulate the catches in future, rather than take more fish. Measure the lengths of all the fish on the sheet to the nearest mm and record your results in a group table using a group size of 0.5 cm. A computer with a statistical or spreadsheet package might be useful in this exercise.

Draw a histogram and a probability distribution curve. Remember that percentage frequencies will need to be divided by the group size (i.e. doubled) for the vertical axis.

What is the probability of catching a fish bigger than 48 cm i.e. $P(x > 4.8)$?
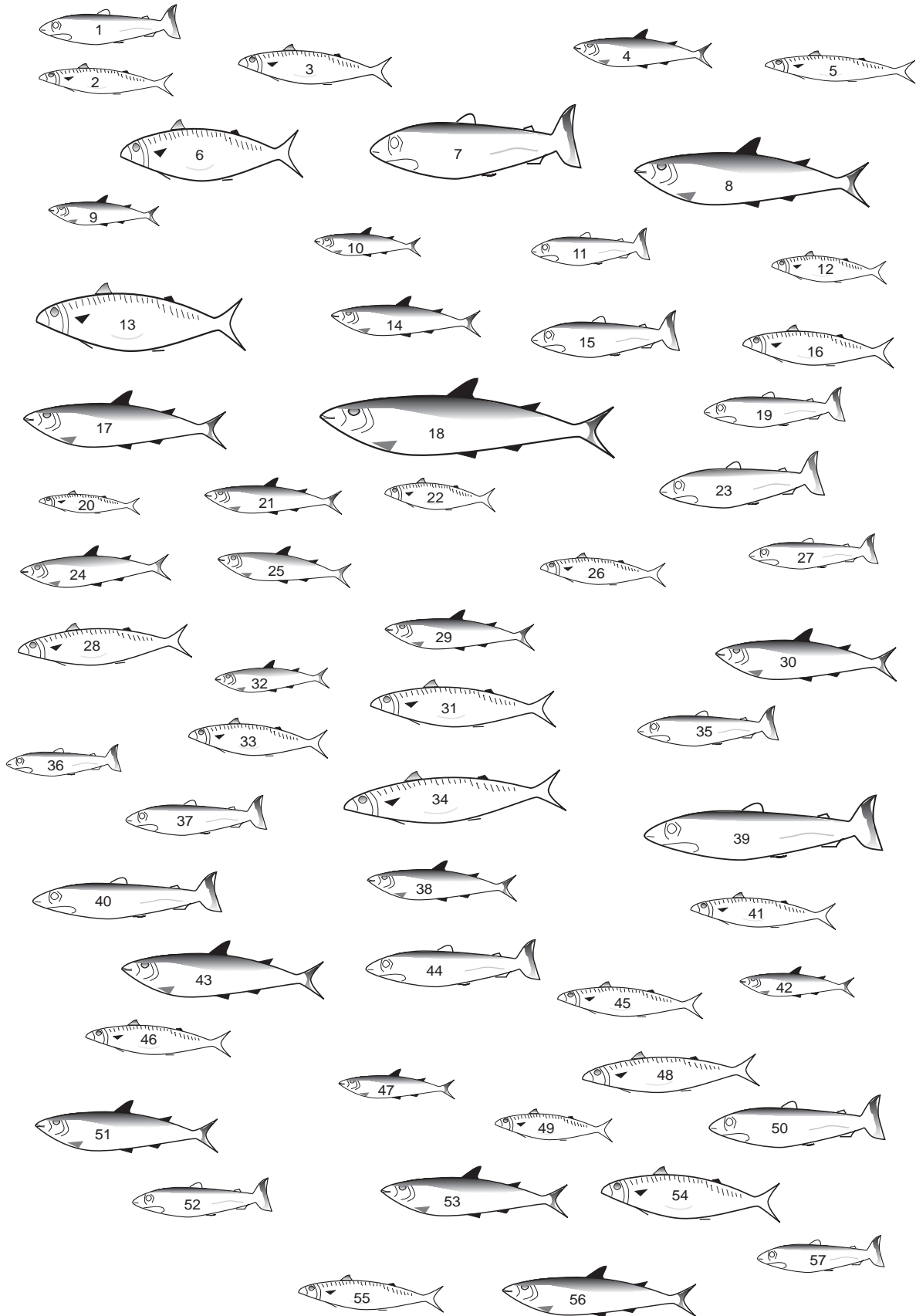
Using a graph is clearly an inefficient and tedious way of calculating probabilities. You should be familiar with the technique of integration, in particular using it to find areas under curves. In order to use this concept, however, it is necessary to find the function to represent the curve.

If you look again at the curve for the shot put throws on page 132, this appears to be an inverted quadratic function. Since it crosses the x-axis at ± 2, these are roots of the equation. The function is therefore of the form

$$f(x) = A\left(4 - x^2\right)$$

where A is constant.

To find the value of A it is necessary to make sure that the total area under the curve between -2 and $+2$ is 1, since the total probability must be one. This can be found by integration as follows:

$$\int_{-2}^{2} A\left(4 - x^2\right) dx \; = \; A\left[4x - \frac{x^3}{3}\right]_{-2}^{2}$$

$$= \; A\left\{\left[8 - \frac{8}{3}\right] - \left[-8 - \frac{-8}{3}\right]\right\}$$

$$= \; \frac{32A}{3}.$$

> Note that questions involving integration will not be set for the AEB examination of Statistics module, so much of this chapter is optional.

This must equal 1, so A must take the value $\dfrac{3}{32}$.

Note that this only works if the range of answers is restricted to -2 to $+2$. This is usually made clear by defining a probability density function (p.d.f.) as follows:

$$f(x) = \begin{cases} \dfrac{3}{32}\left(4 - x^2\right) & \text{for } -2 < x < 2 \\[2mm] 0 & \text{otherwise} \end{cases}$$

Any function which can be used to describe a continuous probability distribution is called a **probability density function**.

*Activity 4   Checking out functions

The scientist in the fish example wants to find a suitable function for her results. What kind of function do you think applies here? Use a graphic calculator or computer package to guess at possible alternatives.

For each of the alternatives below and any of your own, do the following:

(a) Find any constants that are necessary to give a total area of 1 under the curve (assume that all fish are between 0.5 and 5.5).

(b) Sketch the actual curve using a graphic calculator, computer package or table of values over the original histogram. Judge how well the curve fits.

A : quadratic curve
   [Hint: assume it goes through (6, 0)]

B : straight line

C : cosine curve (use radians).

Care must be taken with some functions as, although they may have an area of 1, they may not be functions usable in calculating probability. For example, consider the function illustrated opposite.

Although this function may integrate to give an area of 1, this is because the area below the graph is negative. This is **not** a probability density function, as it would give negative probabilities.

The only other restriction on suitable functions is that the function exists for all values of x in the given range. The best way to check this is to sketch the graph or plot on a graphic calculator.
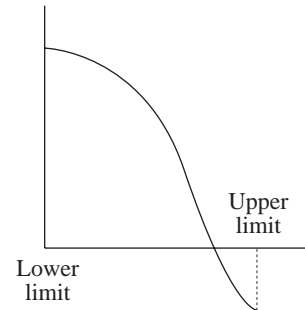
Exercise 7A

1. Check whether the following are suitable probability density functions over the given range:

    (a) $f(x) = \frac{1}{2}(x^2 + 4)$ $0 < x < 1$

    (b) $f(x) = \frac{1}{2}$ $2 < x < 4$

    (c) $f(x) = \frac{x}{4}$ $1 < x < 3$

    (d) $f(x) = \frac{x}{6} + \frac{1}{12}$ $0 < x < 3$

    (e) $f(x) = \frac{1}{2}(2x - 1)$ $0 < x < 2$

2. A variable has a p.d.f. given by

    $$f(x) = A(x^2 + 4) \quad 0 < x < 1.$$

    Find the value of the constant A such that this constitutes a valid p.d.f.

7.3   Calculating probabilities

Once a suitable function has been found, the main purpose of using a p.d.f., that is, to calculate the probabilities of events, can be carried out. In the previous two sections two ideas have been used, namely
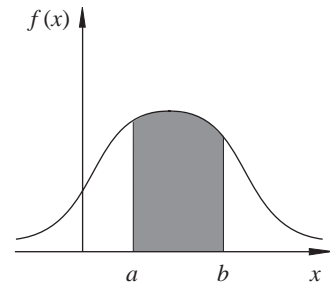
(a)   that the probability of a range of values can be found by finding the area between the values under a p.d.f curve;

(b)   that integration of a p.d.f. can be used to find these areas.

In general, if $f(x)$ is a continuous p.d.f. defined over a specified range of x, then

total area under the curve $=1 \Rightarrow \int_{-\infty}^{\infty} f(x)dx = 1$

and $P(a < x < b)$ is the area under the curve from $x = a$ to $x = b$; this can be written as

$$P(a < x < b) = \int_{a}^{b} f(x)dx$$

*Example

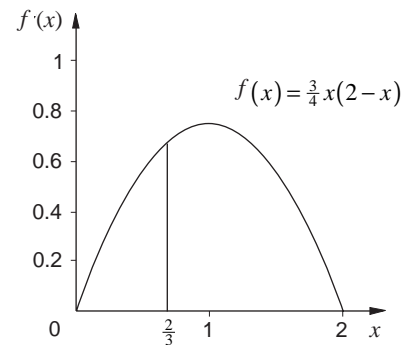The p.d.f. of the age of babies, x years, being brought to a post-natal clinic is given by

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

If 60 babies are brought in on a particular day, how many are expected to be under 8 months old?

Solution

Eight months $= \frac{2}{3}$ year, so

$$P\left(x < \tfrac{2}{3}\right) = \int_{0}^{\frac{2}{3}} \tfrac{3}{4}x(2-x)dx$$

$$= \tfrac{3}{4}\int_{0}^{\frac{2}{3}}\left(2x - x^{2}\right)dx$$

$$= \tfrac{3}{4}\left(x^{2} - \dfrac{x^{3}}{3}\right)_{0}^{\frac{2}{3}}$$

$$= \tfrac{3}{4}\left[\tfrac{4}{9} - \tfrac{8}{81}\right] - [0]$$

$$= \tfrac{3}{4}\left[\tfrac{28}{81}\right] = \tfrac{7}{27} \approx 0.259$$

Hence the expected number of babies under 8 months

$$= 60 \times \frac{7}{27} = 15\frac{5}{9} \approx 15.56.$$

The function which was obtained after integration, and is used to calculate probabilities, is called the **cumulative distribution function**, $F(x)$. In the last example

$$F(x) = \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right), \qquad 0 \le x \le 2$$

It is cumulative, as putting a single value a in the function will give you the probability $P(0 < x < a)$.

For example,

P(baby is less than 6 months old)

$$= F(0.5) = \frac{3}{4}\left(0.5^2 - \frac{(0.5)^3}{3}\right) = \frac{5}{32} = 0.15625$$

Once the cumulative function is known it is unnecessary to repeat the integration for new examples. In general, for a continuous p.d.f., $f(x)$, the cumulative distribution function, is given by

$$\boxed{F(x) = \int_{-\infty}^{x} f(x)dx}$$

*Activity 5   How good are your functions?

In the last activity different functions were suggested to fit the fish data. How can you measure which is the best? For each function find the probability of a value being in each 0.5 cm group. If you find $F(x)$ this should not take long. By multiplying each of these by 57 you can find the 'expected' number in each category according to the function. Make a table showing the expected and observed numbers.

How can an overall measure be found to test which is the best fit?

*Exercise 7B

You may assume, unless stated, that all the functions are valid p.d.f.s. Sketch the function in each case.

1. The resistance of an electrical component follows a p.d.f. given by

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

What is the probability that the resistance is less than 2?

2. A biologist is examining the growth of a virus. A tiny amount is placed on a culture plate and it is found that the surface area in cm^2 occupied by the virus eight hours later is given by the p.d.f.

$$f(x) = \frac{e^x}{19} \quad 0 < x < 3.$$

Check that this is approximately a valid p.d.f. and find the probability that a culture plate has a surface area in excess of $2\ cm^2$.

3. The weekly demand for petrol at a local garage (in thousands of litres) is given by the p.d.f.

$$f(x) = 48\left(x - \tfrac{1}{2}\right)(1 - x) \quad \tfrac{1}{2} < x < 1.$$

The petrol tanks are filled to capacity of 940 litres every Monday. What is the probability the garage runs out of petrol in a particular week?

*7.4 Mean and variance

Activity 6   Response times

Using a stopwatch or a watch with a stopwatch facility in hundredths of a second, set the watch going and try to stop at exactly 5 seconds. Record the exact time on the stopwatch - again a computer facility would help. It is also easier to work in pairs. Repeat this 100 times and draw up a histogram of your results. Find also the mean and variance of your results.

Although it will vary according to how good you are, the following p.d.f. should approximate to your times.

$$f(x) = \begin{cases} \dfrac{375}{32}(5.4 - x)(x - 4.6) & 4.6 < x < 5.4 \\ 0 & \text{otherwise} \end{cases}$$

Sketch this curve and verify that it approximates to your data. Find the expected number of results in each range using the p.d.f. You could then work out the expected mean and variance using these figures, but is there a quicker way?

The idea of the mean and variance of a probability distribution was met in Chapter 4, p89, on discrete distributions. The mean can be found by the same method i.e. $E(X) = \sum x\,P(x)$. This would require multiplying small ranges of x by the area under the curve for the whole defined range.

Clearly this is tedious and to obtain an accurate result would require very small ranges of x. In the last section, integration was used to find areas over a range of values. To find the **mean** of a distribution simply use

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

We often use μ to represent the mean of a distribution. In most cases the function is defined over only a small range so it is not necessary to integrate between $\pm\infty$, only in the defined range. This can be used in the baby example to find the mean age of babies brought to the clinic.

Example

Find the mean age of babies brought to the clinic described in the example in Section 7.3.

Solution

$$
\begin{aligned}
E(X) &= \int_0^2 x \cdot \frac{3}{4} x \cdot (2 - x)\,dx \\[2mm]
&= \frac{3}{4} \int_0^2 \left(2x^2 - x^3\right) dx \\[2mm]
&= \frac{3}{4} \left[\frac{2}{3} x^3 - \frac{x^4}{4} \right]_0^2 \\[2mm]
&= \frac{3}{4} \left[\frac{16}{3} - 4 \right] - [0] \\[2mm]
&= \frac{3}{4} \times \frac{4}{3} = 1.
\end{aligned}
$$

This result should not be surprising since the original sketch showed the distribution to be symmetrical, so the mean must be in the middle of the range. This will always be true, so this could save integrating.

In the same way, the basic definition of **variance** used with discrete distributions can be used, but replacing summation with integration; this gives

$$V(X) = E\left(X^2\right) - \mu^2$$
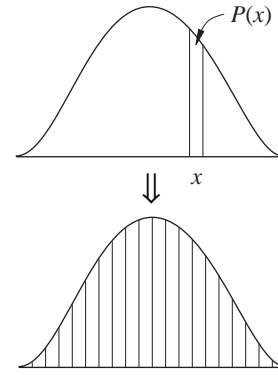
Example

For the babies distribution, find the variance of x.

Solution

$$E(X^2) = \int_0^2 x^2 \frac{3}{4}x(2-x)dx$$

$$= \frac{3}{4}\int_0^2 (2x^3 - x^4)dx$$

$$= \frac{3}{4}\left[\frac{1}{2}x^4 - \frac{1}{5}x^5\right]_0^2$$

$$= \frac{3}{4}\left[8 - \frac{32}{5}\right] - [0]$$

$$= \frac{3}{4} \times \frac{8}{5} = \frac{6}{5}.$$

Hence $\qquad V(X) = \frac{6}{5} - 1^2 = \frac{1}{5}.$

The standard deviation can be found by square-rooting the variance, so for the example above,

$$s = \sqrt{\frac{1}{5}} \approx 0.45.$$

Activity 7

Evaluate the mean and variance of the p.d.f. used in the response times activity. How well do these compare with the actual values?

*Exercise 7C

1. A p.d.f. is given by $f(x)=6x(1-x)$ for $0<x<1$. Find the mean and variance of this distribution.

2. A teacher asks her pupils to draw a circle with some compasses they have been given. The p.d.f. of the radii is given by

$$f(r)=\frac{r}{4} \quad 1<r<3.$$

Find the mean and variance of the radii drawn.

3. The proportion of cloud cover at a particular meteorological office is given by the p.d.f.

$$f(x)=12x(1-x)^2 \quad 0<x<1.$$

Find the mean and variance of this distribution.

4. A p.d.f. is given by $f(x)=ke^{-x}$ for $x>0$. Find the value of k which makes this valid and hence the mean and variance of this distribution.

5. A p.d.f. is given by $f(x)=Ax(6-x)^2$ for $0<x<6$. Find the value of A and hence the mean and variance of this distribution.

*7.5 Mode, median and quartiles

There are, of course, other measures which can be calculated for a p.d.f. Some of these are introduced below.

Mode

The mode is defined as the value which has highest frequency. In a continuous case this is clearly the value of x which gives the maximum value of the function.
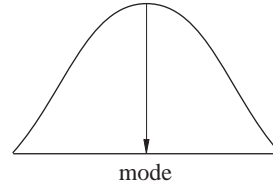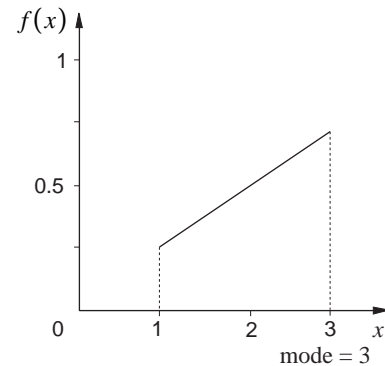

mode

In many cases a simple curve sketch will show this value.

For example, when

$$f(x) = \frac{x}{4} \qquad 1 < x < 3,$$

the mode is clearly at 3.

With more complicated functions it may be necessary to differentiate to find **maxima/minima**.

mode = 3

Example

Find the mode of the p.d.f. defined by

$$f(x) = 12x^2(1-x) \qquad 0 < x < 1.$$
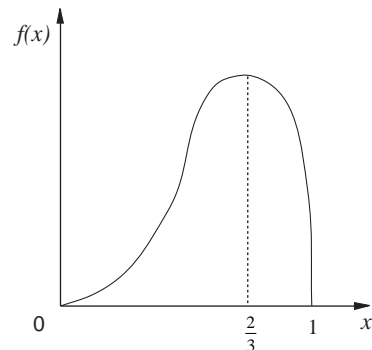
Solution

Since
$$f(x) = 12x^2 - 12x^3$$

$$\frac{d}{dx} f(x) = 24x - 36x^2$$

$$= 12x(2 - 3x)$$

$$= 0$$

when $x = 0$ or $\frac{2}{3}$.

A sketch (or second derivative) would reveal that $x = 0$ is a minimum point and $\frac{2}{3}$ is the mode.

Activity 8

Look at the graphs of p.d.f.s that you drew in the fish example and in some of the exercises. For which cases is the mode obvious and which will require a maxima/minima differentiation method? Are there any for which it would be impossible to find a mode? Sketch some cases which might cause problems.

Median and quartiles

Medians and, in fact, any of the percentiles can be found from their basic definitions. The **median** is defined as the value *m* for which

$$P(x < m) = P(x > m) = \frac{1}{2}.$$

In terms of a continuous p.d.f. this is the value which divides the area into two parts each with area $\frac{1}{2}$.

The value *m* can be found by integration.

Example

Find the median of this p.d.f.

$$f(x) = \frac{x}{4}, \quad 1 < x < 3.$$

Solution

Now
$$\int_1^m \frac{x}{4} dx = \frac{1}{2}$$

$$\Rightarrow \quad \left[\frac{x^2}{8}\right]_1^m = \frac{1}{2}$$

$$\Rightarrow \quad \frac{m^2}{8} - \frac{1}{8} = \frac{1}{2}$$

$$\Rightarrow \quad m^2 - 1 = 4$$

$$\Rightarrow \quad m = \pm\sqrt{5}.$$

Since $-\sqrt{5}$ is outside the range of the function the median must be at $\sqrt{5}$.

144

The same method can be used to find quartiles and other percentiles. However, for higher order polynomial equations solutions can be difficult. It is often simpler to use the cumulative distribution function.

Example

For the babies' weight example used earlier, the cumulative distribution function was given by

$$F(x) = \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right), \quad 0 < x < 2.$$

Show that the median value is $x = 1$, and estimate the interquartile range.

Solution

Now $F(1) = \frac{3}{4}\left(1 - \frac{1}{3}\right) = \frac{3}{4} \times \frac{2}{3} = \frac{1}{2}.$

So the area from $x = 0$ to $x = 1$ is 0.5, and $x = 1$ is the median (as well as the mode and mean!).

For the lower quartile, $F(x) = \frac{1}{4}$,

giving $\frac{1}{4} = \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right)$

$\Rightarrow \quad x^3 - 3x^2 + 1 = 0.$

This has approximate solution $x = 0.65$, and by symmetry the upper quartile will be at $x = 2 - 0.65 = 1.35$.

The inter-quartile range is then given by

$$1.35 - 0.65 \approx 0.70.$$

Exercise 7D

1. Find the mode of these p.d.f.s:

    (a) $f(x) = \frac{3}{50}\left(x^2 - 4x + 5\right) \quad 0 < x < 5$

    (b) $f(x) = \frac{3}{13}\left(x^2 + 4\right) \quad 0 < x < 1.$

2. Find the median of the p.d.f.s given by:

    (a) $f(x) = \frac{1}{8}(4 - x) \quad 0 < x < 4$

    (b) $f(x) = e^{-x} \quad x > 0.$
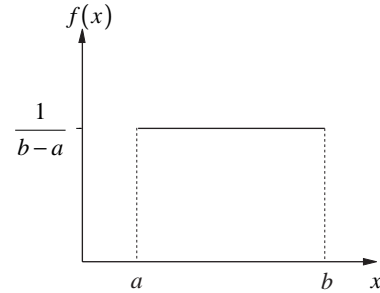
7.6 Rectangular distribution

One very special continuous distribution that does not require calculus to analyse it is the rectangular or uniform distribution. Its p.d.f. is defined as

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

Its shape is illustrated opposite, and it is clear that its mean value is given by

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

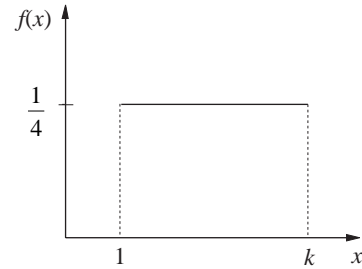i.e. the midpoint of the line between a and b.

* Activity 9

(a) Use integration to verify the formula for $E(X)$.

(b) By integration show that

$$V(X) = \frac{1}{12}(b-a)^2$$

Example

The continuous random variable X has p.d.f. $f(x)$ as shown opposite. Find

(a) the value of k

(b) $P(2.1 < X < 3.4)$

(c) $E(X)$

Solution

(a) The area under the curve must be 1, so

$$\frac{1}{4} \times (k-1) = 1$$

$$\Rightarrow \quad k-1 = 4$$

$$\Rightarrow \quad k = 5$$

(b) Now

$P(2.1 < X < 3.4) =$ area under the curve from $x = 2.1$ to $x = 3.4$

$$= \frac{1}{4} \times (3.4 - 2.1)$$

$$= \frac{1}{4} \times 1.3$$

$$= 0.325$$

(c) $E(X) = \frac{(1+5)}{2} = 3$, using the formula (or by symmetry)

Activity 10   How random are telephone numbers?

Take around a hundred telephone numbers from a local telephone directory at random. Write down the last three digits in each number. These should be evenly spread in the range 000 to 999. (The first digits are often area codes.) Group them using group sizes of 200, i.e. 000-199 etc. and draw a histogram. Find the mean and variance of the data.

If the numbers were truly random and the sample sufficiently large, you would expect the distribution to be rectangular in shape. To form a p.d.f. you need to ensure that the total area of the graph is 1, so with a range of 1000 the p.d.f. is given by
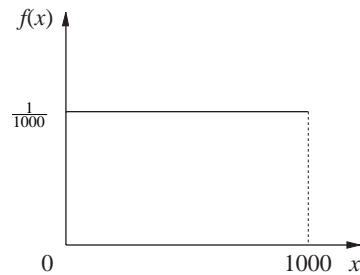
$$f(x) = \frac{1}{1000}, \quad 0 < x < 1000.$$

(In fact, for the data in Activity 10, only integer values are possible, but it is a reasonable approximation to use a continuous p.d.f.)

So the probability that a randomly chosen telephone number has the last three digits less than 300 is given by

$$\frac{300}{1000} = 0.3.$$

By symmetry the mean is given by $E(X) = 500$.

Using the formula, the variance is given by

$$V(x) = \frac{1}{12}(1000 - 0)^2$$

$$= \frac{1000\,000}{12}$$

$$= \frac{250\,000}{3}$$

$$\approx 83333$$

giving a standard deviation of

$$s \quad \approx \quad 289.$$

Check these against the values you obtained in your telephone survey.

*7.7 Miscellaneous Exercises

1. The distribution of petrol consumption at a garage is given by

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

where x is in thousands of litres. Find the values of a and b if the mean consumption is 600 litres. Hence find the probability that in a given week the consumption exceeds 900 litres.

(AEB)

2. A p.d.f is given by

$$f(x) = kx^2(3-x) \text{ for } 0 < x < 3.$$

Calculate

(a) the value of k to make this valid,

(b) the mean, μ, and variance, s^2, of the distribution and verify that $\mu + 2s = 3$,

(c) the probability that x differs from the mean by more than $2s$.

3. A continuous random variable, X, has p.d.f

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

(a) Determine the value of a.

(b) Sketch the p.d.f.

(c) Find the value of $E(X)$ and $P(X < E(X))$.

What does this tell you about the median of the distribution?

4. A meat wholesaler sells remnants of meat in 5 kg bags. The amount in kg of inedibles (i.e. bone and gristle) is a random variable, X, with p.d.f.

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

(a) Show that $k = \frac{3}{4}$.

(b) Find the mean and variance of X.

(c) Find the probability that X is greater than 2.5 kg. (AEB)

5. A small shopkeeper sells paraffin. She finds that during the winter the daily demand in gallons, X, may be regarded as a random variable with probability density function

$$f(x) = \begin{cases} kx^2(10-x) & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that $k = 0.0012$.

(b) Find the mean and the standard deviation of the distribution.

(c) Find the value of x which makes $f(x)$ a maximum. What is this value called?

(d) Estimate the median using the approximate relationship

$$2(\text{median} - \text{mean}) = \text{mode} - \text{median}.$$

Verify that your answer is approximately correct by finding the probability that an observation is less than your estimate of the median.

(e) If the shopkeeper has storage facilities for only eight gallons and can only replenish her stock once a day before the shop opens, find her mean daily sales.   (AEB)

6. A teacher travels to work by car and the journey time, t hours, has a probability density function

$$f(t) = \begin{cases} 10ct^2 & 0 \le t < 0.6 \\ 9c(1-t) & 0.6 \le t \le 1.0 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

(a) Find the value of c and sketch the graph of this distribution.

(b) Write down the most likely journey time taken by the teacher.

(c) Find the probability that the journey time will be

(i) more that 48 minutes;

(ii) between 24 and 48 minutes.
                                               (AEB)

7. In a competition with a crossbow, contestants aim at a target with radius 5 cm. The target has a bull in the middle of it, of 2 cm radius. Hitting the bull scores 5 points and the outer circle 2. The p.d.f. of the variable X, the distance of a randomly fired shot from the centre of the target, is given by

$$f(x) = \begin{cases} 0.25e^{-0.25x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find

(a) the probability of hitting the bull,

(b) the probability of missing altogether, and

(c) hence the expected score of a single shot.

8. A p.d.f. is given by

$$f(x) = \frac{2}{3}\cos\left(x - \frac{\pi}{6}\right) \quad 0 < x < \frac{2\pi}{3}.$$

(a) Show that this is a valid p.d.f.

(b) Find the mode of the distribution and hence sketch the curve.

(c) Find the probability that $X < 1$.

9. The life of an electronic component is given by the p.d.f.

$$f(x) = \frac{100}{x^2} \text{ hours for } x > 100.$$

Find

(a) the median life of a component.

(b) the probability that a component lasts for more than 250 hours.